



universität
wien

MAGISTERARBEIT

Titel der Magisterarbeit:

„Heuristiken, Intuition und Big Data -
Entscheidungstheorien im Zeitalter komplexer
Datenmengen“

Verfasser:

Philipp Ganster (Bakk. phil.)

Angestrebter akademischer Grad:

Magister der Philosophie (Mag. phil.)

Wien, 2015

Studienkennzahl lt. Studienblatt: 066 841

Studienrichtung lt. Studienblatt: Publizistik und Kommunikationswissenschaft

Betreuer: Univ.-Prof. Dr. Jürgen Grimm

Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe oder Nutzung anderer als der angegebenen Nachweise verfasst habe. Die aus fremden Quellen wörtlich oder inhaltlich entnommenen Stellen sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht. Die vorliegende Fassung ist identisch mit der eingereichten elektronischen Version.

Wien, April 2015

Philipp Ganster

Wien, am 19. 04. 2015

Danksagung

Mein besonderer Dank geht an erster Stelle an Professor Dr. Jürgen Grimm, der mich in der Konzeptionsphase dieser Arbeit, deren Zustandekommen in der Vorgeschichte nicht einfach und von einem durch einen Abgang vom Institut berührten Betreuerwechsel geprägt war, betreute. Auch in der Diskussion über die theoretische Operationalisierung des qualitativen Zugangs erwies sich mir Dr. Grimm jederzeit als Ansprechperson, die mein Erkenntnisbestreben unterstützte und in seiner Betreuung viel Zeit aufbrachte.

Mein Dank geht ebenso an Freunde und die Familie, die mir über diesen Prozess unterstützend und motivierend zur Seite standen.

Philipp Ganster

Inhaltsverzeichnis

Ehrenwörtliche Erklärung.....	3
Danksagung.....	4
1. Einleitung	7
2. Theoretische Operationalisierung.....	10
3. Aktueller Stellenwert.....	12
3.1 Persönliche Motivation.....	12
3.1.1 Zur Person Mayer-Schönberger	17
3.2 Das Spannungsfeld	18
3.3 Die größere gesellschaftliche Perspektive.....	22
4. Begriffsdefinition	25
5. Big Data und der Entscheidungsprozess.....	32
6. Eigenheiten in den Limitierungen von Big Data	42
6.1 Google Flu Trends.....	43
6.2 Aus dem Kontext.....	44
7. Der unsichtbare Algorithmus	45
8. Wissenschaftstheorie	47
8.1 Korrelation contra Kausalität	47
8.2 Das Wissenschaftsverständnis nach Karl Popper.....	50
8.3 Das Wissenschaftsverständnis nach Kuhn	53
8.4 Betrachtung der theoretischen Kompatibilität.....	58
9. Urteilsheuristiken	62
9.1 Kahneman.....	62
9.2 Gigerenzer	65
10. Ethik.....	69
10.1 Bewusste Partizipation und Wahrung der Privatsphäre.....	69
10.2 Datenschutz, ein Ausblick.....	71
10.3 Verteilung von monetärem Wert.....	73

10.4 Das soziale Selbstbild.....	75
10.5 Recht auf Vergessenwerden.....	78
10.6 Anonymisiert und nicht anonym.....	80
11. Fazit.....	83
Quellenverzeichnis	87
Onlinequellen	92
Abbildungsverzeichnis.....	97
Wissenschaftlicher Lebenslauf	98
Abstract.....	100

1. Einleitung

Wir sprechen dieser Tage vermehrt von Daten. Es vergeht kaum ein Tag in der dieser ominöse, fast staubtrockene Begriff nicht Teil der täglichen Berichterstattung wird. Ursächlich ist eine Entwicklung in unserer Gesellschaft, die so umfänglich ist, dass sie vom Ansatz einer neuen Bürgerbeteiligung, über die Etablierung neuer Märkte, bis hinein in die tiefsten Ebenen unserer Privatsphäre reicht – gerade dann, wenn Daten in großen Mengen anfallen schärft sich unser Blick auf sie.

Big Data ist die Geschichte eines angekündigten Paradigmenwechsels. Eine Revolution des quantitativen Elements bei der die Betrachtung individueller Positionen in Koexistenz mit einer beinahe unlimitiert generierbaren Anzahl von Datenpunkten geschieht, die in Aussagen zu Trends und Zusammenhängen zusammenlaufen. Zum Anderen eine Neuerung im Zugang bei dem nicht mehr zwangsläufig eine konkrete Frage forschungsanleitend zeichnet und Hypothesen nicht ausschließlich durch den Vorsatz eines tieferen Verständnis der Materie, sondern ebenso durch wahrgenommene Stärken in der Gerichtetheit von Zusammenhängen entstehen.

Big Data bricht auch mit dem Dogma des betrachtbaren Rahmens der nicht bereits zu Beginn eines Forschungsprozesses abgesteckt wird, sondern individuell als Teil des Analyseprozesses verändert werden kann und sich von mehreren Jahren bis hin zur aktuellsten Ausprägung in Echtzeit erstreckt. In besonderen Fällen werden dadurch Feedbackloops möglich, die den Betrachter selbst zum Akteur werden lassen, oder es ihm erlauben vor dem Hintergrund aktueller Ereignisse Voraussagen anzustellen.

Big Data ist zudem eine Herausforderung für den qualitativen Forschungsansatz innerhalb der Sozialwissenschaften, da sich durch die unmittelbare Testbarkeit von Hypothesen deren inhärenter Charakter und Anspruch verändert. Insbesondere, wenn Aussagen über Verhaltenszusammenhänge getroffen werden sollen, wie sie in

den Geistes- und Sozialwissenschaften Forschungsgegenstand sind – ist die Veränderung eines Prozesses auch diesbezüglich zu reflektieren.

Auch das Adressatenverständnis von Public Relations, oder das Bevölkerungsverständnis in der Politik werden zunehmend von größeren Datenmengen beeinflusst, die über automatisierte Feedbackkanäle anfallen.

Populäre, wenn auch weitläufig vertretene, Leitsätze wie „Sex sells“ reichen zur Umschreibung eines Adressatenverständnisses mittlerweile ebenso wenig aus wie die Frage danach zu stellen, über welche Kanäle Personen am Effizientesten erreicht wurden.

Der moderne Mensch ist in seinem Mediennutzungsverhalten flexibel, nutzt bei gleichbleibenden Aufmerksamkeitsspannen Medienangebote parallel, selektiert Informationen bei sinkenden Schwellenniveaus schneller als je zuvor und bedient sich mittelbar verfügbarer, sozial vernetzter Meinungen um Entscheidungen hinsichtlich seines Verhaltens – womit sich auch das Entscheidungsverhalten größerer Personengruppen in der Gesellschaft verändert. Und an jedem einzelnen Schnittpunkt dieser Entwicklung fallen Daten an.

In komplementären Forschungsrichtungen bilden sich in jüngerer Zeit vermehrt Betrachtungs-Ansätze heraus, denen ein stärkerer Adressaten-Fokus zu Grunde liegt, der über die Betrachtung des Individuums als unscharf definiertem, reagierenden Rezipienten hinausgeht und nach den Ursachen für Entscheidungsverhalten fragt. Diese Ansätze werden auch von Vertretern der Big Data Analyse referenziert um Entscheidungsverhalten begründbar zu machen.

Ziel dieser Arbeit ist es die Veränderungen an Entscheidungsprozessen vor dem Hintergrund einer zunehmenden Datenvielfalt zu beleuchten. Dies geschieht unter Aufarbeitung sozialwissenschaftlicher und sozialökonomischer Theorien die die Themenkomplexe Big Data und Entscheidungsfindung berühren. Insbesondere soll dabei der neo-behavioristischen Begriff der „Black Box“ in dem Entscheidungsprozesse als von außen nicht einsehbar definiert sind, hinterfragt und bestehende, sowie neue Ansatzmöglichkeiten für die Forschung in einen Bezug zueinander gesetzt

werden, um sie hinsichtlich ihres Stellenwerts im wissenschaftlichen Kanon, sowie ihrer Aussagewirkung zu analysieren. Ein weiterer Stellenwert liegt in dieser Arbeit auf der Analyse der theoretischen Kompatibilität der unterschiedlichen Modelle.

2. Theoretische Operationalisierung

Dieser Arbeit liegt der Vorsatz zu Grunde, den aktuellen Stand der Vereinbarkeit der Big Data Analyse mit dem vorherrschenden wissenschaftlichen Paradigma der Sozialwissenschaften, das vor allem als deduktionsgeleitet verstanden wird, zu vergleichen und zu verstehen.

Zu diesem Zweck wurde die Methode der Sekundäranalyse in der Form einer vergleichenden Literaturanalyse gewählt.

Die Sekundäranalyse verlangt ein vorab definiertes Kriterienschema bei der Wahl der analysierten Quellen, da diese in der Folge als Primärquellen das Datenmaterial eines empiriegestützten Zugangs ersetzen.

Da es sich bei der Annäherung der Forschungsrichtungen um ein aktuell andauerndes Unterfangen handelt, bekommt die Aktualität des Primärmaterials, anhand dessen die theoretische Vereinbarkeit geprüft wird, eine besondere Bedeutung. Wo möglich, wurde daher auf Publikationen aus den Jahren 2013 bis 2015 zurückgegriffen und das verstärkt international und im englischen Sprachraum, da nicht erwartet werden kann, dass sich alle Ansätze bereits in ein breiteres Verständnis innerhalb der nationalen Disziplinen durchgeschlagen haben.

Ein weiteres Interesse dieser Arbeit gilt den möglichen gesellschaftlichen Implikationen der steigenden Relevanz von Big Data Anwendungen, weshalb auch dezidiert nach Vertretern einer Meinungslinie innerhalb des aktuellen gesellschaftspolitischen Prozesses gesucht wurde. Es wurde dabei angenommen, dass steuerungspolitische Entscheidungen, die auf diesem Gebiet momentan fallen, die Ebene der gesellschaftlichen Auswirkungen am deutlichsten prägen werden.

Um diesbezüglich keiner eingebundenen Gewichtung aufzusitzen und ein möglichst breites Verständnis des Themenfeldes zu erhalten, wurden explizit auch die Reakti-

on von Meinungsgegnern und im gesellschaftlichen Rahmen, von Vertretern aus dem Bereich der Counterculture einbezogen.

Auch Detailveränderungen einzelner Positionen über den Beobachtungszeitraum hinweg, so von diesen in den Forschungsrichtungen gezeigt, werden betrachtet.

Zur Operationalisierung des Vergleichs der theoretischen Vereinbarkeit, wurden die beiden post-positivistischen Wissenschaftstheorien herangezogen die das aktuelle Verständnis von dem was Wissenschaft zu erfüllen habe, seit den 60er Jahren dieses Jahrhunderts am deutlichsten prägen.

Neuere Betrachtungsansätze in der Wissenschaftstheorie und bezüglich des Entscheidungsprozesses wurden nach einer Revision des Methodendesigns nur noch dort berücksichtigt, wo sie im aktuellen Diskurs als Referenzgrößen fallen, um zwar eine Erweiterbarkeit des aktuellen Wissenschaftsverständnisses zuzulassen, dieses aber nur dort zu bemühen, wo es von den Proponenten in den Primärquellen eingefordert wird.

Diese Arbeit beansprucht für sich eine Momentaufnahme zu sein die versucht den aktuellen Stand des interdisziplinären Diskurses zu verstehen und mögliche zukünftige Entwicklungen in ihren Ansätzen zu beschreiben.

3. Aktueller Stellenwert

3.1 Persönliche Motivation

Im Oktober 2013 hält im Atrium des Leopold Museums in Wien, im Rahmen einer vom Forum Alpbach getragenen Vortragsreihe¹, der Rechts- und Politikwissenschaftler Viktor Mayer-Schönberger einen Frontalvortrag um die Ideen hinter seiner neuesten Veröffentlichung, „Big Data: Die Revolution, die unser Leben verändern wird“ zu bewerben. Den geneigten Zuhörer, der den Text bereits in Vorbereitung auf die anschließende Fragerunde gelesen hat, erwarten neben vielen Einzelfallbetrachtungen - zur Illustration - die er bereits aus der Lektüre des Textes kennt, einige Prämissen, die vor dem Hintergrund der üblichen Entscheidungswelten eines Bildungsbürgers disruptiv und stellenweise nicht wenig provokativ anmuten.

Es geht dabei, knapp skizziert, um die Automatisierbarkeit von Gesellschaftsprozessen, um neue Ansätze zur Voraussagbarkeit von Entwicklungen, um Selektoren die weder öffentlich werden müssen, noch in einer Einzelbetrachtung rechtfertigbar sind und wie man diesen Prozess als Chance begreifen kann – wie in ihm aktuell gearbeitet und gedacht werden kann, sowie welche neuen relevanten Gesellschaftsperspektiven sich daraus entwickeln lassen.

Mayer-Schönberger proklamiert an diesem Abend unter anderem wiederholt die Relevanz eines „Big-Data Anwalts“, zugänglich für jedes Individuum einer Gesellschaft, dessen Aufgabenfeld analog zu einer neuen Version einer Konsumentenberatungsstelle darin angesetzt werden müsse, Personen im Konfliktfall jene Entscheidungsprozesse zu erläutern die, von ihnen diffus wahrgenommen, gerade einen Aspekt ihres Leben beeinflusst haben, und ihre Auslegung im Interesse des Individuums zu prüfen.

¹ Alpbach Talks (2013) – Themenreihe „Digitale Welten“: Der digitale Mensch im Web 3.0, Montag, 7. Oktober 2013, 19.00 – 20.30 Uhr, Museumsquartier, Wien;
<http://www.alpbach.org/de/unterveranstaltung/alpbach-talks-big-data/>, [Letzter Aufruf: 23. 02. 2015;]

Es ist an dem Punkt als - paraphrasiert - vom Vortragenden die Aussage „Big Data ist eine Möglichkeit Finanzkrisen wie die von 2008 zu verhindern“ fällt, als die Stimmung im Saal kippt. „Wir hatten nicht zu viele Daten, wir hatten zu wenig um das System zu verstehen“ ist ein Ausspruch den einige Teilnehmer im Saal nicht unbedingt gutieren.

Ich frage Mayer-Schönberger an diesem Abend, ob auch die zur öffentlichen Kenntnis gelangten Enthüllungsberichte über die Aggregations- und Auswertungspraxis großer, sensibler Datenmengen durch den US Geheimdienst für Inneres – NSA, in der Person von Edward Snowden², die zum Zeitpunkt des Vortrags vier Monate alt sind, und die öffentliche Berichterstattung darüber, etwas an seinem Zugang, oder Verständnis der Materie geändert haben. Mayer-Schönberger verneint dies, mit einem Verweis auf sein Buch in dem die Thematik bereits angerissen sei.

Ich erwähne als Teil meiner Fragestellung die Person von Caspar Bowden der von 2002 bis 2011 als leitender Datenschutzbeauftragter (Chief Privacy Adviser) von Microsoft Corp. tätig war und aus dem Unternehmen im letzten Jahr dieser Periode ausgeschieden ist, nachdem er begonnen hatte auf Unvereinbarkeiten mit seinem Aufgabenfeld hinzuweisen.

According to Microsoft's then-Chief Privacy Advisor, Caspar Bowden, he warned dozens of colleagues in 2011, two full years before Snowden's National Security Agency disclosures became worldwide news. Specifically, he says, he told dozens of colleagues that increasingly gutted American privacy laws, thanks to the 2008 FISA Amendment Act, meant the NSA could conduct "unlimited surveillance" on cloud computing data sold to foreign countries. [...]

² „June 5, 2013: First revelations arising from the documents provided by Snowden are published in a Guardian article about the NSA's collection of domestic email and telephone metadata from Verizon as part of what is later revealed to be an even broader collection effort.“ Williams, Brian (2014). NBC News, <http://www.nbcnews.com/feature/edward-snowden-interview/edward-snowden-timeline-n114871>, [Letzter Aufruf: 23. 02. 2015];

He says he was promptly threatened to be fired and then, two months later, he was let go “without cause” after a nine-year stint at the company. (O’Neill, 2015)

Schönberger gutiert das Fallbeispiel und gibt an es zu kennen, geht darüber hinaus in der Fragebeantwortung jedoch nicht näher darauf ein.

Ich stelle mir an diesem Abend die Frage, inwieweit der kritische Umgang mit Reaktionen auf Auswertungen in der Big Data Analyse bereits Teil des wissenschaftlichen Prozesses geworden ist und inwieweit Gesellschaftsprozesse generell auf das Selbstverständnis der Akteure Einfluss nehmen können – selbst wenn einige davon so deutlich wahrnehmbar sind, dass sie monatelang Berichterstattung an sich binden.

Mayer-Schönbergers Verweis auf die Betrachtung im Rahmen seines Buches, der mein Interesse an dem Themengebiet bestärkt hat, lässt sich prüfen. Der Text ist auch in digitaler Form verfügbar³ und so bietet sich als Erstes die Stichwortsuche innerhalb des Volltextes an.

Auf den 257 Seiten (Umfang der Printausgabe) ist der Begriff „Snowden“ kein einziges Mal enthalten –

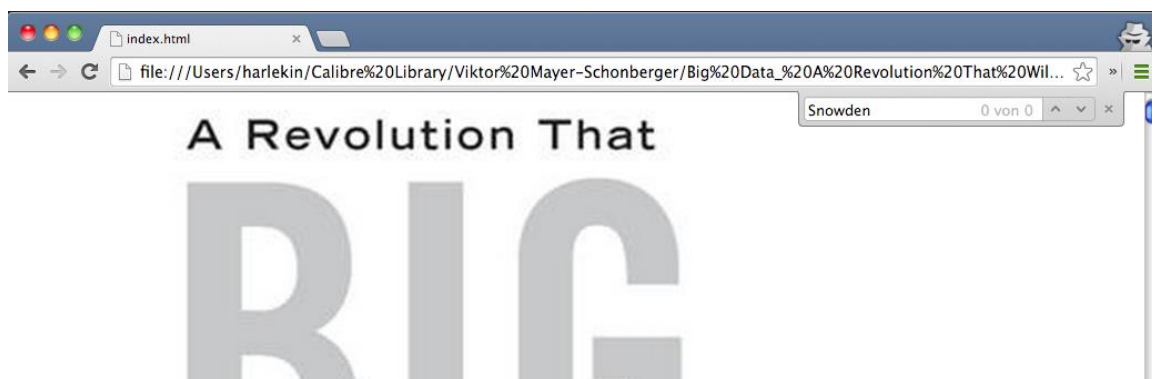


Abbildung 1 – Volltextsuche nach „Snowden“

³ Viktor Mayer-Schönberger, Kenneth Cukier - Big Data: A Revolution That Will Transform How We Live, Work, and Think [Kindle Edition], 2013, Amazon Digital Services, Inc.;

- der Begriff NSA findet sich im Text vier Mal, davon zwei Mal als Teil des Quellenverzeichnisses – sodass die beiden aufeinanderfolgenden Paragraphen die gesamte Kontextualisierung beinhalten, hier zitiert werden können.

[...] the U.S. National Security Agency (NSA) is said to intercept and store 1.7 billion emails, phone calls, and other communications every day, according to a Washington Post investigation in 2010. William Binney, a former NSA official, estimates that the government has compiled “20 trillion transactions” among U.S. citizens and others—who calls whom, emails whom, wires money to whom, and so on [Anm. Eine Umschreibung für die Praxis Sammeln von Metadaten]. To make sense of all the data, the United States is building giant data centers such as a \$1.2 billion NSA facility in Fort Williams, Utah. And all parts of government are demanding more information than before, not just secretive agencies involved in counterterrorism. When the collection expands to information like financial transactions, health records, and Facebook status updates, the quantity being gleaned is unthinkable large. The government can’t process so much data. So why collect it?

The answer points to the way surveillance has changed in the era of big data. In the past, investigators attached alligator clips to telephone wires to learn as much as they could about a suspect. What mattered was to drill down and get to know that individual. The modern approach is different. In the spirit of Google or Facebook, the new thinking is that people are the sum of their social relationships, online interactions, and connections with content. (Mayer-Schönberger 2013, Kapitel 8, zweiter Absatz, ff;)

In der Lektüre dieser Absätze eröffnet sich das Selbstverständnis eines neuen Feldes.

Während der Text vordergründig auf einen – laut Eigendefinition neuen Ansatz eingeht, der darin bestehe Individuen durch ihre Sozialgraphen zu beurteilen, gibt es an dieser Stelle kaum theoriebezogene Einschränkungen. Daten müssten aus dem Prozessdenken heraus auch dann gesammelt werden, wenn diese keinen eindeutigen

Anlass zur Analyse erkennen ließen, da in ihnen nachfolgend Zusammenhänge für Gruppenverhalten und das Verhalten einzelner zu finden seien, die zu einem späteren Zeitpunkt immer noch von Bedeutung sein könnten.

Einzelne Aspekte einer solchen Auswertung werden im weiteren Verlauf des Werkes durchaus kritisch hervorgehoben -

The trouble is that humans are primed to see the world through the lens of cause and effect. Thus big data is under constant threat of being abused for causal purposes, of being tied to rosy visions of how much more effective our judgment, our human decision-making of assigning culpability, could be if we only were armed with big-data predictions. (Mayer-Schönberger 2013, Kapitel 8, Unterkapitel: Probability and punishment, erster Absatz;)

Mayer-Schönberger sieht die häufige Attribution des Kausalitätsprinzips, auf dessen Bedeutung innerhalb des Big Data Paradigmas in dieser Arbeit noch nachfolgend eingegangen wird, als eines der Hauptprobleme des Forschungsansatzes – akzeptiert jedoch gleichzeitig die im Einzelfall anlasslose Aggregation und Speicherung von Massendaten unter Vorbehalten wie Prävention.

In many contexts, data analysis is already employed in the name of prevention. [...] Individuals with certain characteristics are subjected to extra screening when they pass through airport security.

That's the idea behind "profiling" in today's small-data world. Find a common association in the data, define a group of people to whom it applies, and then place those people under additional scrutiny. It is a generalizable rule that applies to everyone in the group. "Profiling," of course, is a loaded word, and the method has serious downsides. If misused, it can lead not only to discrimination against certain groups but also to "guilt by association." [...]

Most important, using big data we hope to identify specific individuals rather than groups; this liberates us from profiling's shortcoming of making every predicted suspect a case of guilt by association. In a big-data world, somebody with an Arabic name, who has paid in cash for a one-way ticket in first class, may no longer be

subjected to secondary screening at an airport if other data specific to him make it very unlikely that he's a terrorist. With big data we can escape the straitjacket of group identities, and replace them with much more granular predictions for each individual. (Mayer-Schönberger 2013, Kapitel 8, Unterkapitel: Probability and punishment, zwölfter Absatz;)

Einen Vorteil dieses Ansatzes sieht Mayer-Schönberger darin, gruppenbezogene Vorurteile abzubauen und durch ein techniziteres Verständnis eines neuen Gruppenbegriffs zu ersetzen zu können.

Nach einer ersten genaueren Betrachtung des Problemfeldes war es für mich anleitend von besonderem Interesse zu verstehen, wie eine Forschungsrichtung gleichzeitig scheinbar dem Kausalzusammenhang abschwören kann, und sich trotzdem als vielleicht wichtigste Quelle gesellschaftlichen Fortschritts ihres Jahrzehnts sieht.

3.1.1 Zur Person Mayer-Schönberger

Viktor Mayer-Schönberger, Professor am Oxford Internet Institute, ist einleitend nicht beliebig gewählt, sondern kann als ein Hauptproponent hinter dem 2011 von der Europäischen Kommission im Rahmen der angestrebten EU Datenschutzreform aufgegriffenen „Recht auf Vergessenwerden“ genannt werden, das den Big Data Komplex ebenfalls berührt und das aktuell in Form der Richtlinie 95/46/EG eine Entscheidung des EuGH in einem Verfahren der Europäischen Union gegen Google Inc. maßgeblich beförderte.

Nachdem die erwähnte EuGH Entscheidung öffentliche Kritik nach sich zog, nahm der EuGH dazu im September 2014 Stellung–

EuGH-Vizepräsident Koen Lenaerts erläuterte im Interview mit der taz[9], dass der Gerichtshof kein „Recht auf Vergessenwerden“ erfunden habe. Er habe nur eine Interessenabwägung auf Grundlage der EU-Datenschutz-Richtlinie vorgenommen. ⁴ (Wikipedia, Eintrag „Recht_auf_Vergessenwerden“, – Bezug nehmend auf: Das Recht auf Privatheit überwiegt (2014), taz.de⁵)

Am 6. Februar 2015 veröffentlichte Google Inc. die Ergebnisse eines Expertenbeirats, darunter auch der Wikipedia Gründer Jimmy Wales, der sich uneins über eine Umsetzung der diesbezüglichen Entscheidung zeigte.⁶ Als Hauptkritikpunkt kristallisiert sich heraus, dass das Recht der Öffentlichkeit auf Information vor diesem Hintergrund möglicherweise zu wenig Betrachtung fand, und von kritischen Proponenten als „Geschichtsrevisionismus unter dem Vorwand von zur Anwendung gebrachten Persönlichkeitsrechten“ zum Vorwurf gemacht wird. Eine Abwägungsfrage.

Im Mai 2014 veröffentlichte Google Inc. nichtsdestotrotz das Frontent mit dem Bürger der europäischen Union eine Löschung von Sucheinträgen beantragen können.⁷

3.2 Das Spannungsfeld

Am 21. 08. 2014 stellt das europäische Forum Alpbach, in einer inhaltlichen Fortsetzung, eine Podiumsdiskussion zum Thema „Big Data: Ein Bergwerk ohne Regeln?“, zu der erstmalig auch der erwähnte Casper Bowden geladen ist.⁸

Der Vortrag erhält eine Einleitung von Alexander Klimburg, der an der Harvard Kennedy School Cambridge zur inneren und äußeren Sicherheitspolitik der Europä-

⁴ Recht auf Vergessenwerden, http://de.wikipedia.org/wiki/Recht_auf_Vergessenwerden, Datum des letzten Aufrufs: 24. 02. 2015;

⁵ Rath, Christin (2014). Das Recht auf Privatheit überwiegt. <http://www.taz.de/1/archiv/?dig=2014/09/19/a0084>, [Letzter Aufruf: 24. 02. 2015];

⁶ APA (2015), "Lösch-Beirat" von Google uneins über Recht auf Vergessen, <http://futurezone.at/netzpolitik/loesch-beirat-von-google-uneins-ueber-recht-auf-vergessen/112.272.815>, [Letzter Aufruf: 24. 02. 2015];

⁷ siehe: FAZ (2014), So können Google-Nutzer Suchergebnisse löschen lassen, <http://www.faz.net/aktuell/wirtschaft/netzwirtschaft/neues-formular-im-internet-so-koennen-google-nutzer-suchergebnisse-loeschen-lassen-12964383.html> [Letzter Aufruf: 11.04.2015]

⁸ Die Veranstaltung ist als Audiovortrag in ihrer vollen Länge auf der Webseite des Forums unter <http://www.alpbach.org/de/vortrag/big-data-ein-bergwerk-ohne-regeln/> abrufbar.

ischen Union forscht. Klimburg war unter anderem in beratender Funktion für die österreichische Delegation der Organization for Security and Co-operation in Europe (OSCE) tätig.⁹

Klimburg beginnt seine Einleitung mit folgendem Statement:

„The bilingual speakers among you might already have noted that a literal translation of the German title would be „Big Data as a mine“. This translation was not used, because the obvious connotation of mine is usually the military mine, the one that blows things up, like people and vehicles. But strangely enough that would be oddly appropriate for this session, because we already had our first casualty. Elizabeth Linder of facebook [Anm.: Politics and Government Specialist for the EMEA Regions, facebook, London] decided, that the subject of Big Data was too dangerous of a minefield for her to cross, so she unfortunately dropped out – but that means that we have more time with Casper Bowden, and we can also talk about something that he is more interested in – and I am more interested in, and that is not only Big Data, but Big Data and surveillance.“

„As the other English title of our session implies, data is the fuel that drives our present data age. During the late 19th century the industrial revolution was largely powered by coal. In fact one of the key aspects of industrialisation was that coal could be used to produce more coal – which was used in a variety of other steam engines from factories to ships. Data today seems to have a similar function, powering internet services, that would otherwise cost money and help develop new products and services. But the comparisons do not stop there. The unprohibited use of coal in the 19th century had considerable externalities [...], with serious implications for public health for example. While the misuse of data has probably not killed anyone yet, many observers are concerned [what] the large amount of data can reveal about individuals and how the information can be used. (Klimburg, 2014)

Bereits die Einleitung eröffnet, in welchem Spannungsfeld unterschiedliche Proponenten des wissenschaftlichen Diskurses auch in diesem Zusammenhang stehen.

⁹ vgl. ebd.

Einerseits gilt Big Data als die Triebfeder des aktuellen Informationstechnologiezeitalters – andererseits sei „misuse“ an der Tagesordnung, und die Aufbruchsstimmung insbesondere durch die Annahme einer uneingeschränkten Nutzungsperspektive bedingt.

Das Fernbleiben des Industrievertreters wird an dieser Stelle genutzt um vermehrt den Überwachungsaspekt zur öffentlichen Diskussion zu bringen, auf den sich der Vortrag in Folge auch beinahe ausschließlich bezieht. Da der Aspekt in der Einleitung dieser Arbeit bereits einmal angerissen wurde, entfällt an dieser Stelle eine erneute Detailbetrachtung.

Zur weiteren Kontextualisierung des Fernbleibens der Facebook Sprecherin sei erwähnt, dass Facebook Inc. am 24. Februar diesen Jahres, nach einem von der Europäischen Kommission beauftragten Forschungsbericht (vgl. Brendan et al., 2015) massiv durch den EU-Kommissar Günther Oettinger (Commissioner for Digital Economy & Society) bezüglich der Unvereinbarkeit ihrer Geschäftsbedingungen, sowie ihrer handlungsleitenden Praxis mit der geltenden Europäischen Rechtsauffassung kritisiert wurde.¹⁰

Der Forschungsbericht der Universität Leuven attestiert Überschreitungen in acht juristischen Feldern, darunter EU Vertragsrecht, Datenschutz- und Persönlichkeitsrecht.

Am 02. April 2015 veröffentlicht die Forschungsgruppe der KU Leuven in Zusammenarbeit mit der Universität Brüssel einen Folgebericht (Annex 1)¹¹ in dem zusätzlich auf Verstöße im Rahmen der social tracking Aktivitäten (insbesondere dem Tracking von Nicht-Nutzern) von Facebook hingewiesen wird, was eine erneute Welle an Medienberichten nach sich zieht.¹²

¹⁰ vgl. Fairless, T. (2015) Europe's Digital Czar Slams Google, Facebook. <http://www.wsj.com/articles/europes-digital-czar-slams-google-facebook-over-selling-personal-data-1424789664>, [Letzter Aufruf: 24. 02. 2015];

¹¹ vgl. <http://www.law.kuleuven.be/icri/en/news/item/icri-cir-advises-belgian-privacy-commission-in-facebook-investigation>, [Letzter Aufruf: 11. 04. 2015];

¹² vgl. Gibbs S. (2015) <http://www.theguardian.com/technology/2015/feb/23/facebooks-privacy-policy-breaches-european-law-report-finds> [Letzter Aufruf: 11. 04. 2015];

In einer Veröffentlichung vom 08. April 2015 spricht ein Unternehmenssprecher in diesem Zusammenhang von einer technischen Fehlfunktion, die von Facebook Inc. schnellstmöglich behoben werde.¹³

Im Mai 2014 spricht Mayer-Schönberger im Rahmen einer Frontaldiskussion auf der re:publica 2014¹⁴, einem jährlichen Vernetzungstreffen der deutschen „digitalen Gesellschaft“, zum Thema Big Data, diesmal bereits mit einem anderen Focus. Der Vortrag trägt den Titel „Freiheit und Vorhersage – über die ethischen Grenzen von Big Data“. Mayer-Schönberger eröffnet hier mit einem Fallbeispiel das die Problematik von Verhaltensvorhersage durch Big Data Anwendungen verdeutlicht. Weiters spricht er über Grenzen von Vorhersagen die laut seinem Dafürhalten rechtlich ausschließen müssen, dass der freie Wille von Individuen im Sinne von Gesellschaftsentscheidungen übergangen wird (im Verständnis einer Vorhersage) und somit erstmalig über ein explizites Verbot einer speziellen Form von Big Data basierter Urteilsfindung.

So könne ein Individuum durchaus dazu motiviert werden eine vorhersagbare Handlung (im Beispiel ein Verbrechen) nicht zu begehen, es müsse jedoch ausgeschlossen werden, dass es aufgrund einer hohen Wahrscheinlichkeit ein Handlung zu begehen bestraft werden würde. Anderenfalls käme das einer Aufhebung des humanistischen Konzepts des freien Willens gleich. Mayer-Schönberger beschreibt nachfolgend die Gedanken- und Entscheidungsfreiheit als notwendige Ventile in einer Gesellschaft die aufgrund gesammelter Daten „nicht mehr vergisst“

Für den Fall dass der freie Wille nicht ausgesetzt, sondern nur durch das Setzen von Verhaltensanreizen oder die Reduktion von Entscheidungsmöglichkeiten, limitiert wird, verweist Meyer-Schönberger erneut auf die gesellschaftliche Relevanz eines

¹³ vgl. Biselli A. (2015) <https://netzpolitik.org/2015/its-a-bug-not-a-feature-sagt-facebook-zum-tracking-von-nicht-usern/> [Letzter Aufruf: 11. 04. 2015];

¹⁴ In seiner Gesamtlänge online auf der Webseite des Veranstalters unter <http://republica.de/file/republica-2014-viktor-mayer-schoenberger-freiheit-un> einsehbar. [Letzter Aufruf: 19.04.2015]

„Big-Data Anwalts“, diesmal jedoch explizit auch darauf, dass eine Möglichkeit zur Falsifikation von Big Data basierten Richtlinien, für die Einzelperson, gegeben sein muss, und mehr noch – dass bestimmte ein Individuum betreffende Aussagen und Charakterisierungen von vorne herein durch einen etablierten Prozess zertifiziert sein müssten, sodass zumindest ein Teil des Aufwands zur Sicherstellung von Validität bis hin zum Einzelfall auf der Seite des Big Data Nutzers (im Gegensatz zum einzelnen Subjekt) liegt. (vgl. Mayer-Schönberger, 2014)

3.3 Die größere gesellschaftliche Perspektive

Alexander Klimburg nimmt in dem weiter oben genannten Vortrag Bezug auf eine zeitnahe Enquete mit der er in der Vorbereitung auf den Vortrag in Kontakt gekommen sei, und die die gesamtgesellschaftlichen Erwartungen die auf den Methodenkomplex projiziert werden unterstreiche. Es geht an dieser Stelle um den als besonders sensibel bezeichneten Bereich des (österreichischen) Gesundheitssystems und wie dessen weitere Entwicklung von mehreren Seiten antizipiert wird.

In den im August 2014 präsentierten Ergebnissen der „Gesundheitsgespräche 2014¹⁵“ unter dem den Rahmen steckenden Untertitel „Gesundheitswesen im Jahr 2025“, bilden sich erneut gewisse Vorstellungen ab.

Drei Arbeitskreise umreißen hier kurz das Feld in dem sie tätig wurden, gefolgt von einer Präsentation der „wichtigsten erarbeiteten Punkte“.

Der Generaldirektor des Hauptverbandes der Sozialversicherungsträger, Josef Probst, erklärte im Arbeitskreis "Big Data als Chance für ein bürgerorientiertes Gesundheitssystem 2025", dass es im Jahr 2025 eine mit den Bürgern entwickelte europäische Strategie für "Big Data" inklusive einer breiten und öffentlichen Forschungsstrategie in Sachen Gesundheit geben werde. Öffentlich finanzierte Forschungsprojekte und deren Ergebnisse müssten aber auch öffentlich nutzbar bleiben und dürften nicht allein privaten Patentinhabern überantwortet werden, so Probst.

¹⁵ vgl. (APA, 2014)

[...]

Im Arbeitskreis "Big Data – Selbstbestimmtes Gesundheitsverhalten des Einzelnen oder geschickte Manipulation?" zeichnete Marcus Müllner, Geschäftsführer der PERI Change GmbH, das Bild, dass es im Jahr 2025 nicht mehr um "Big Data", sondern um "Smart Data" [Anm.: impliziert gesetzte Kausalbedeutungen] gehen werde. Bis dahin seien einerseits die Ziele für eine erfolgreiche Gesundheitsförderung erfüllt, andererseits die Standards, die notwendige Datenqualität für Analysen und die "Deutungshoheit" – wer interpretiert die individuellen oder öffentlichen Gesundheitsdaten – geklärt.

[...]

Jan Oliver Huber, Generalsekretär des Verbandes der pharmazeutischen Industrie (Pharmig), leitete den Arbeitskreis "Bremst der Datenschutz die Weiterentwicklung unseres Gesundheitssystems?". Huber erklärte, dass auf dem Weg zu einer Wissensgesellschaft im Gesundheitswesen noch erhebliche Stolpersteine lägen: "Es fehlt an politischem Wille. Wir haben in Österreich schon jetzt sehr viele Daten in öffentlichen Institutionen, die nicht zur Verfügung gestellt werden." (APA, 2014)

Und erneut findet sich eine Beraterfunktion (hier in einer definitiven Form) unter den Forderungen.

Angesichts der enormen Herausforderungen durch Themen wie Big Data, Neurowissenschaften und Genetik forderten viele ExpertInnen die Einrichtung eines Weisenrates (z.B. um Vorschläge zur Nutzung von Patientendaten für Forschungszwecke zu entwickeln etc.) (APA, 2014)

Unter Betrachtung der aktuellen Relevanz lässt sich hier bereits festhalten, dass die Frage zu Entscheidungsprozessen in diesem Umfeld eine sehr vordergründige und

dringliche ist und dass wiederholt eine Erwartungshaltung besteht diese durch Expertengremien geklärt zu sehen.

Big Data wird umfassend, von Akteuren der öffentlichen Hand, der Privatwirtschaft und aus dem wissenschaftlichen Umfeld, als eine der nächsten gesellschaftlichen Entwicklungsperspektiven gesehen.

Zudem ist der zeitliche Rahmen zur Umsetzung, beziehungsweise die Erwartungshaltung wann diese geregelt in weiten Bereichen des öffentlichen Lebens erfolgen kann, durchaus knapp bemessen.

Der wissenschaftsgeleitete „Big Data Experte“ der gesellschaftsrelevante Fragestellungen mit seiner Expertise anleitend begleiten und Implementationsschritte bewerten soll, wird sowohl von den Vertretern der Theorie, als auch von Personen im Umfeld der Umsetzung nachgefragt. Dabei geht es explizit nicht um die Bewertung von technischen Problemfällen, sondern um Hilfestellungen zu Fragen der sozialen Vertretbarkeit, Prozessrisiken und Ethik.

4. Begriffsdefinition

Nach der obenstehenden, freien Nutzung des Begriffs „Big Data“ innerhalb seiner gesellschaftlichen Konnotation, wird es notwendig den Begriff hinsichtlich seines Bedeutungsgehalts innerhalb der wissenschaftlichen Disziplinen und seiner Verwendung innerhalb dieser Arbeit zu definieren.

Der britische Ökonom und Mitglied des Nuffield College der Universität Oxford¹⁶, Tim Harford - für Royal Dutch Shell und die Weltbank in beratender Funktion tätig, charakterisiert in einem Artikel in der Financial Times¹⁷ Mayer-Schönbergers Definition von Big Data wie folgt:

Professor Viktor Mayer-Schönberger of Oxford's Internet Institute, co-author of Big Data, told me that his favoured definition of a big data set is one where "N = All" – where we no longer have to sample, but we have the entire background population. Returning officers do not estimate an election result with a representative tally: they count the votes – all the votes. And when "N = All" there is indeed no issue of sampling bias because the sample includes everyone. (Harford, 2014)

Dieser sehr holistischen Auffassung des Begriffs – die im Übrigen auch ein Licht darauf wirft wie Mayer-Schönberger dem von ihm geprägten Begriff der „Zwangsjacke der Gruppenidentitäten“ zu entfliehen sucht, sei gleich eine etwas mehr technokratische entgegengesetzt.

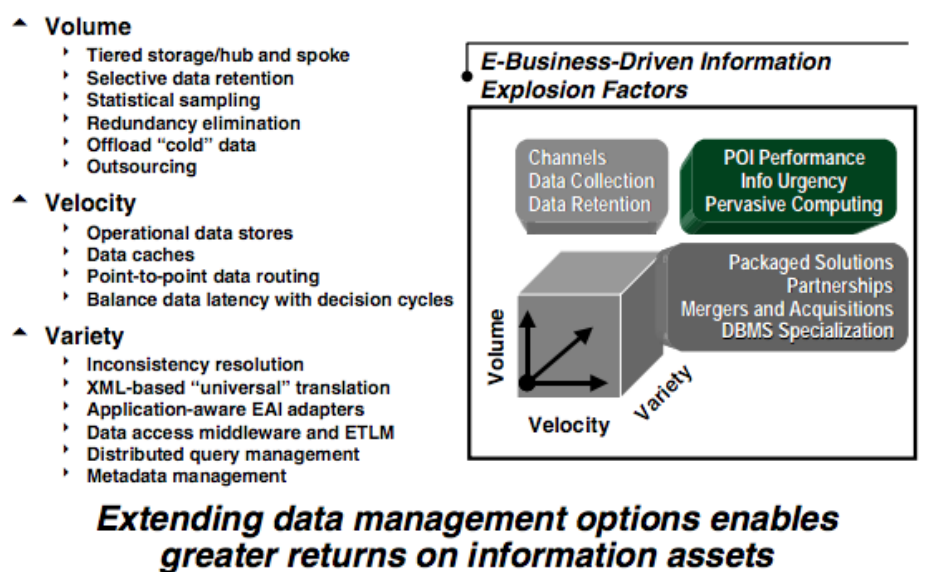
So prägt Douglas Laney (VP, Information Innovation and Strategy, des Thinktanks Gartner Group) den Begriff in einer mit hoher Resonanz belegten Publikation von 2012 wie folgt:

¹⁶ vgl. <http://timharford.com/etc/biography/>, [Letzter Aufruf: 26. 02. 2015];

¹⁷ Harford (2014) Big data: are we making a big mistake?
<http://www.ft.com/intl/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html>, [Letzter Aufruf: 26. 02. 2015];

Big data is high volume, high velocity, and[or] high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization (Laney, 2012)

Diese Definition geht auf ein von Laney selbst 2001 veröffentlichtes Schema¹⁸ zurück, das den drei genannten Kategorien weitere Unter Aspekte zuordnet die den Umgang mit Datenpunkten und die strukturelle Zusammensetzung des gesamten Aggregats beschreiben.



Source: META Group

Abbildung 2 – Big Data Metrik: Volume, Velocity, Variety, nach Laney

Hier die drei Hauptkategorien im Überblick.

Das „Volume Management“ beschreibt eine Staffeung bei der Speicherung von Daten, sowie – für die Verwendung in den Anwendungsbereichen notwendig, das selektive Vorhalten von Daten, ihre Fassung in statistischen Samples, die Eliminierung von Mehrfacherfassungen, das Auslagern „kalter“ Datenbestände sowie deren nachfolgendes outsourcing.

¹⁸ vgl. Laney (2001) 3-D Data Management: Controlling Data Volume, Velocity and Variety

Eine generelle Löschung, oder eine automatische (Aus-)Selektion ausgehend von einem gewissen Alter der Bestände ist nach dieser immer noch vorhaltenden Definition kein Teil des „Volume Managements“. Der Aspekt soll an dieser Stelle hervorgehoben werden.

Das „Management der Fließgeschwindigkeit“ (Velocity) beschreibt die Zusammenführung mehrerer Datenbestände die auf unterschiedliche Speicherorte verteilt sind. Sie ist nach prozesstechnischen Aspekten angelegt, die sich als technische Kriterien in der Verarbeitung zusammenfassen lassen. Dazu zählt insbesondere die Optimierung der Durchsatzrate, die bedingt wie schnell auf verteilt gelagerte Information zugegriffen werden kann.

„Big Data“ beschreibt somit Datenmengen die nicht mehr zentral auf einem Rechner gespeichert werden müssen, sondern über Cluster verteilt in der Nutzung zusammengeführt werden. Notwendig wird das durch die schiere Menge an Information die zunehmend gesammelt werden kann. Wenn wir gesellschaftspolitisch von einer neuen „Qualität“ des Daten-Sammelns sprechen, findet sich in „Velocity“ die technische Entsprechung zur Operationalisierung dieses Gedankens.

Das „Management der Vielfalt“ (Variety) ergibt sich aus der Granularität mit der die Datenbestände analysiert werden sollen (Inconsistency resolution), einer Übersetzung der Inhalte in eine einheitliche Datenstruktur (XML-based „universal“ translation) zur nachfolgenden Verarbeitung, dem Schnittstellendesign (Application-aware EAI adapters), der technischen Implementierung (Drittanbieter Software beim Datenzugriff, sowie im eigentlichen Funktionsprozess (Extract, Transform, Load and Management)), der Notwendigkeit zur parallelen Abarbeitung von Anfragen (queries) und dessen Strukturierung, sowie aus der Verwaltung von Metadaten, für die weitere Klassifikation und Bedeutungszuweisung.

Wir sehen an dieser Stelle ein ausschließlich prozessorientiertes Verständnis bei der Begriffsdefinition vorherrschend, das sich aus dem Umstand ergibt, wie große Datenmengen, operational handhabbar, aggregiert und verarbeitet werden können. Weitere Aspekte, die beispielsweise umschreiben, was aus der Verarbeitung solcher

Datenmengen hervorgeht, oder auf welchen erkenntnisorientierten Grundannahmen diese fußt, finden wir an dieser Stelle nicht in der Begriffsdefinition.

Eine generelle Definition nach einem Schwellenwert bei der Datenmenge, ist nach Bill Franks, einem Fakultätsmitglied des International Institute for Analytics, ebenfalls nicht möglich, zudem sei eine solche generell schwierig da –

What's big data today won't be considered big data tomorrow any more than what was considered big a decade ago is considered big today. Big data will continue to evolve.[...]

As discussed at the start of the chapter, the accepted definitions of big data are somewhat "squishy." There is no specific, universal definition in terms of what qualifies as big data. Rather, big data is defined in relative terms tied to available technology and resources. As a result, what counts as big data to one company or industry may not count as big data to another. [...]

More important, what qualifies as big data will necessarily change over time as the tools and techniques to handle it evolve alongside raw storage size and processing power. (Franks 2012, S. 24)

Wir sehen an dieser Stelle sogar einen expliziten Verweis auf die selbst wissenschaftsintern „unscharf“ abgrenzbare Definition des Big Data Begriffs der in seiner verbreitetsten Verwendung auf technische Kriterien bei der Verarbeitung fußt.

Franks repliziert ebenfalls auf die Schwierigkeiten bei der Selektion, und in der Idee Datensätze überhaupt zu verwerfen.

There will be pieces that have long - term strategic use, pieces that have short - term tactical use, and pieces that don't matter at all. It will seem odd to let a lot of data slip past, but that is par for the course [...]. Throwing data away will take some getting used to. (Franks 2012, S. 19)

Ein kurzer Verweis auf die Größe bereits operationalisiert nutzbarer Datensammlungen, findet sich bei Harford – und soll an dieser Stelle nur einen kurzen Eindruck die möglichen Größenordnungen vermitteln.

Some emphasise the sheer scale of the data sets that now exist – the Large Hadron Collider’s computers, for example, store 15 petabytes a year of data, equivalent to about 15,000 years’ worth of your favourite music. [...] Such data sets can be even bigger than the LHC data – Facebook’s is – but just as noteworthy is the fact that they are cheap to collect relative to their size, they are a messy collage of datapoints collected for disparate purposes and they can be updated in real time. (Harford, 2014)

In dieser Definition finden sich erneut die kaum definierte Intention im Vorfeld einer Aggregation, sowie ein Verweis auf die potentielle Echtzeitkomponente in der Verarbeitung und Analyse.

Ein letzter an dieser Stelle hervorgehobener Aspekt der obenstehend in Form einer Nennung der oftmals vielschichtigen Nutzungsambitionen solcher Datensätze Anklang findet ist, dass diese vermehrt an verschiedenen Schnittstellen bereits in einer größeren Menge anfallen und eine Aggregation daher kein interessen- oder zielvorstellungsgeleiteter Prozess sein muss, sondern kosteneffizient bereits im Hinblick auf einen vermuteten Zusatznutzen erfolgen kann.

Was die Anwendung von Big Data Applikationen in kleineren Unternehmen angeht, empfiehlt Franks in solchen Fällen eine Limitierung der Erhebungsperiode und des Blickfeldes, keines Falls jedoch eine konkrete Limitierung im Hinblick auf ein zuvor definiertes Interesse.

[...], instead of setting up formal processes to capture, process, and analyze all the data all the time, capture some of the data in a one-off fashion. Perhaps a month’s worth for one division for a certain subset.

[...]

At this point, it is time to turn analytic professionals loose on the data. Remember: they're used to dealing with raw data in an unfriendly format. They can zero in on what they need and ignore the rest. They can create test and control groups to whom they can send the follow-up offers, and then they can help analyse the results. During this process, they'll also learn an awful lot about the data and how to make use of it. (Franks 2012b, S. 18)

Franks weist an dieser Stelle darauf hin, dass sich selbst den Analysten der Sinn oder Nutzen einer Erhebung erst nach dem Durcharbeiten des Datenmaterials eröffnen würde – außerdem wird dieser Prozess als erfahrungsgeleitet umschrieben, die Nützlichkeit der Daten erschlossen sich erst einem erfahrenen Betrachter.

Dies spricht vor allem im Bezug auf das vorherrschende Selbstverständnis unter Big Data Analysten.

Abschließend bleibt festzuhalten, dass Big Data in der aktuellen Wortbedeutung nicht durch äußere Kriterien wie Umfang oder eine konkrete zur Anwendung gebrachte Methodologie in der Auswertung definiert wird – die im Hinblick auf diese Arbeit von besonderem Interesse wären, sondern durch bestimmte Eigenheiten die dadurch festgesetzt sind dass ab einer bestimmten Größe (Schwellenwert verschiebt sich ständig) des Datensatzes neue Verarbeitungsmethoden notwendig werden um mit diesem erfolgreich zu arbeiten. Eine weitere Konkretisierung wird von einigen Proponenten des Forschungsansatzes dezidiert ausgeschlossen, da die technische Entwicklung jede „zu enge“ Definition fortwährend brechen würde.

In der weiteren Folge dieser Arbeit sollen unter dem Begriff „Big Data“ vor allem Anwendungsfelder betrachtet werden, bei denen die Datenmenge als umfangreich genug gesehen wird um im Hinblick auf die Entwicklung von Gesellschaftsprozessen Aussagen zu treffen.

Die Blickrichtung ist vor allem unter Einbeziehung der folgende Problemlage gewählt:

A preliminary examination of the debates, discussions, and writings on Big Data demonstrates a pronounced lack of consensus about the definition, scope, and character of what falls within the purview of Big Data. A meeting of the Organization for Economic Cooperation (OECD) in February of 2013, for instance, revealed that all 150 delegates had heard of the term "Big Data," but a mere 10% were prepared to offer a definition, perhaps worryingly, these delegates "were government officials who will be called upon to devise policies on supporting or regulating Big Data" (Ekbja et al. 2014, S. 3)

5. Big Data und der Entscheidungsprozess

Unhandlich zu fassen und dennoch stark verbreitet als Objekt des aktuellen wissenschaftlichen Diskurses.

Ein möglicher Betrachtungsansatz sich die gesteigerte Popularität von Big Data als Untersuchungsgegenstand innerhalb der Wissenschaft zu erklären, liegt in der Veränderung von Gesellschaftsprozessen durch das Internet.

Im Februar 2014 erscheint im Journal für Behavioral and Brain Sciences Vol. 37 der Cambridge University Press, eine Sammlung an wissenschaftlichen Artikeln rund um eine interdisziplinären Forschungsarbeit unter der Ideenführerschaft von R. Alexander Bentley, mit dem Titel „Mapping collective behavior in the big-data era“.

In ihr wird folgendes wissenschaftstheoretisches Problem skizziert:

The behavioral sciences have flourished by studying how traditional and/or rational behavior has been governed throughout most of human history by relatively well informed individual and social learning.

In the online age, however, social phenomena can occur with unprecedented scale and unpredictability, and individuals have access to social connections never before possible. Similarly, behavioral scientists now have access to “big data” sets – those from Twitter and Facebook, for example – that did not exist a few years ago. (Bentley et al. 2014, S. 63)

Es geht im Teil darum, dass aktuelle soziale Phänomene von den Verhaltenswissenschaften partiell nicht mehr angemessen beschrieben werden können, da sich sowohl der Informationszugang (Schwellendynamik), als auch die Einflussosphäre einer wichtigen sozialen Orientierungsgröße (des Social Web) geändert haben. Die Wissenschaft sieht in Big Data analog dazu eine Möglichkeit Aussagen über zunehmend komplexe Vorgänge treffen zu können.

Weiters wird angemerkt, dass bestimmte Datensätze für Wissenschaftler erstmalig, in dieser Größenordnung zur Verfügung stehen. Auf Anfrage, oder durch Eigenaggregation, die nun viel stärker automatisierbar und dadurch verteilt durchführbar ausfällt. Ein weiteres Interesse ergebe sich daher bereits aus der Möglichkeit des Zugangs.

Der Artikel beschreibt im Weiteren einen Prozess der es erlaubt Datensätze, die Entscheidungsmuster von untersuchten Populationen abbilden, einem behavioristischen Entscheidungsmodell zuzuordnen, indem einzig das Ausprägungsmuster ihrer Verteilung (in der Gestalt einer Kurve, siehe Abb. 3) in einem von vier Quadranten verortet wird, welcher den Entscheidungsstrukturen unterschiedliche Entscheidungsbedingungen zuordnet. (vgl. Abb. 4)

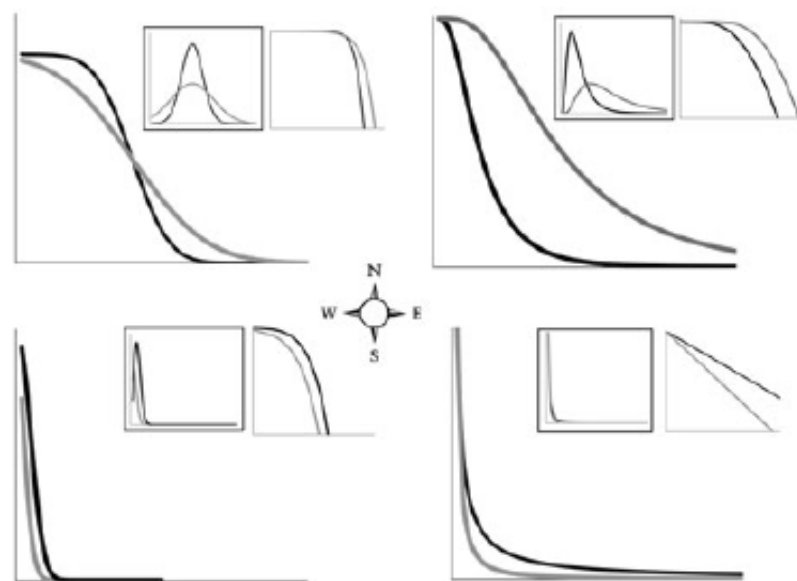


Abbildung 3 - „Generalisierte Verteilungen die die unterschiedlichen Quadranten der Karte charakterisieren. Normal (Gauss) im Nordwesten, negativ binominal im Südwesten, log-normal im Nordosten und power-law im Südosten“ nach Bentley et al. (2014).

Die beiden Achsen in Abb. 3 bilden dabei, im Konzept der Autoren, durch eine von diesen selbst getroffene Charakterisierung - ab, ob Entscheidungen in Populationen „unabhängig, oder in starker sozialer Wechselwirkung zu einander getroffen werden“, und ob die Entscheidungsoptionen, sowie ihre Auswirkungen von Populationen „distinkt und klar erkennbar, oder unklar und dispers wahrgenommen werden“

(kann als Informiertheit umschrieben werden, wobei den Autoren nach, auch Komplexität und Vielfalt auftretende Muster gegen „Süden“ transformieren würde.)

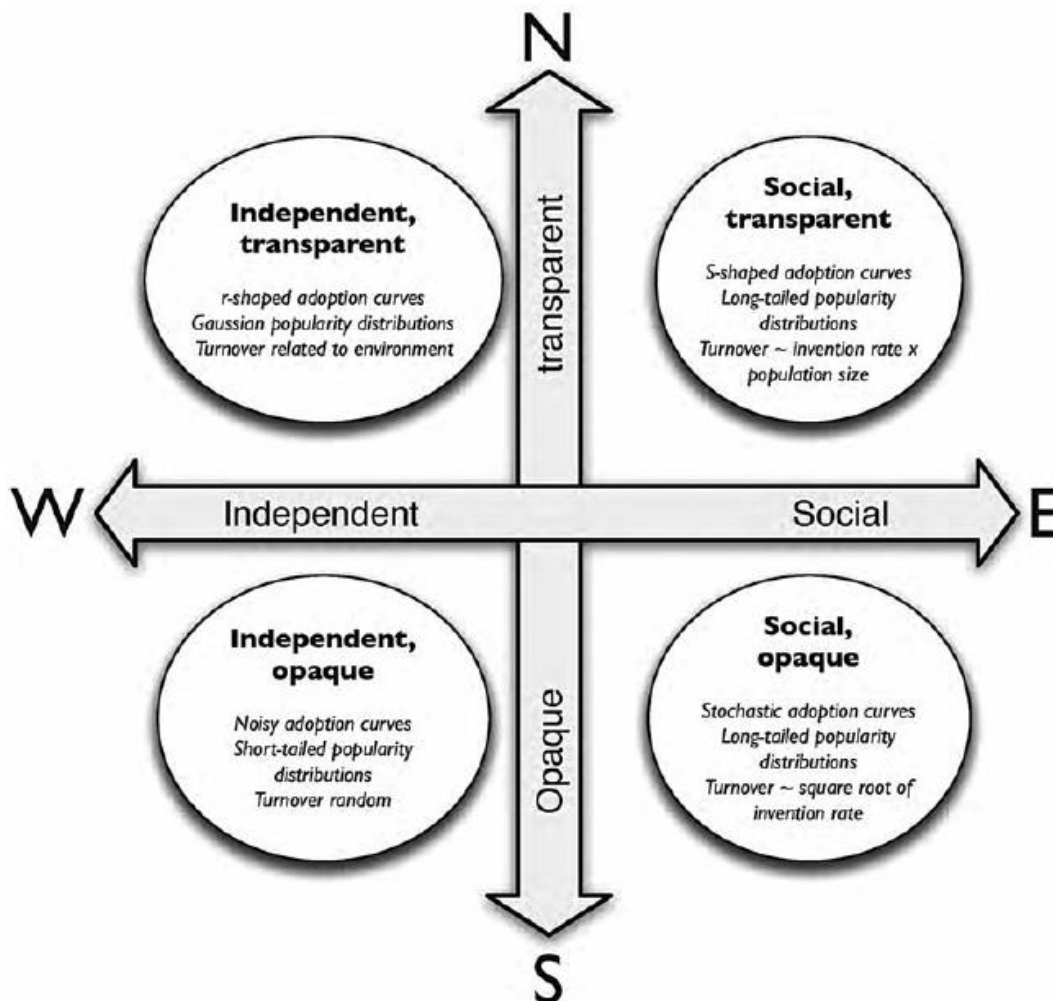


Abbildung 4 - „Zusammenfassung der Vier-Quadranten-Karte zum Verständnis unterschiedlicher Felder des Entscheidungsprozesses von Personen, darauf basierend ob eine Entscheidung unabhängig, oder sozial beeinflusst getroffen wird, sowie der Transparenz der Entscheidungsoptionen sowie ihres jeweiligen „Payoffs“ nach Bentley et al. (2014).

Verschiebt sich ein bestimmtes Entscheidungsmuster über den Zeitverlauf in seiner Verortung innerhalb der Quadranten, so ginge das einher mit einer Veränderung im Entscheidungsverhalten der untersuchten Population. Dies wäre von besonderem Interesse, da sich aufgrund des zunehmenden Echtzeitcharakters der erhebbaren Daten (die einfache statistische Auswertung kann automatisiert geschehen) Entwicklungsänderungen auch antizipieren lassen würden, beispielsweise wenn Datenpunkte innerhalb eines Quadranten zum jeweiligen Extrem tendieren – vor allem aber auch um einen long Tail an bestehenden Daten die über einen vergangenen

Zeitraum aggregiert wurden, rückblickend einer Auswertung unterziehen zu können, die Rückschlüsse auf den Verlauf vergangener Entscheidungs-Perioden erlauben würde, die mit bestimmten Ereignissen verknüpft werden können, was einen Einblick in die Entscheidungsmuster dahinter erleichtere. Somit könnten auch vergangene Ereignisse schneller und besser interpretiert werden.

Die Verortung in einem der vier oben gesehenen Quadranten würde, so Bentley et al., im Anschluss bereits selbsttätig, unterschiedliche wissenschaftliche Untersuchungsrichtungen nahelegen

the northwest (economic), the southeast (cultural-historical), the northeast (both), and the southwest (neither). (Bentley et al. 2014, S. 74)

und weiters anleitend dabei helfen Hypothesen bezüglich einer möglichen Kausalität zu formen.

Es handelt sich hier somit um einen Versuch aus dem gesammelten Auftreten von sozialen Entscheidungsmustern die Art der Entscheidungsfindung in der beobachteten Population abzulesen und zu gewichten.

Die gesellschaftstheoretischen Ambitionen dieses vereinheitlichten Frameworks zur Auswertung von Massendaten fallen dabei nicht gering aus. Der folgende, schwer aufzugliedernde Paragraph beschreibt dies eindrücklich.

The degree of social influence on decision making is an empirical question that underlies what big data means and how they[sic!] can be used. As an example of the importance of this issue, consider the ubiquitous reliance on crowdsourcing in behavioral studies, business, and politics (Horton et al. 2011) – what Wegner (1995) termed “transactive memory” and now commonly called the “wisdom of crowds” (Couzin et al. 2011; Lorenz et al. 2011; Surowiecki 2004): Ask a question of a group of diverse, independent people, and the errors in their answers statistically cancel, yielding useful information. Wikipedia is founded on this assumption, of course, even

though copying of text is essential to its growth (Masucci et al. 2011). The wisdom-of-crowds effect is lost, however, if agents are not thinking independently (Bentley & O'Brien 2011; Salganik et al. 2006). There are numerous indications that online behavior may be getting more herdlike, more confused, or even more "stupid" (Carr 2008; Onnela & Reed-Tsochas 2010; Sparrow et al. 2011). In economies replete with online communication and a constant barrage of information – often to the point of overload (Hemp 2009) – crucial human decision making might be becoming more herdlike in contexts such as voting (Arawatari 2009) and forming opinions about climate change (Ingram & Stern 2007), mating (Lenton et al. 2008; 2009), music (Salganik et al. 2006), and finances (Allen & Wilson 2003). (Bentley et al. 2014, S. 64)

Bentley et al. versuchen über diesen Ansatz das Funktionsmodell hinter Konzepten wie Schwarmintelligenz oder dem attestierbarem Auftreten von Herdenverhalten in der Entscheidungsfindung von Populationen zu entschlüsseln welches, so oben genannt, auch im Prozessverlauf demokratischer Wahlen zu beobachten ist.

Dabei ergehen auch Überlegungen in eine verhaltenskorrektive Richtung von Gesellschaftsprozessen, beziehungsweise in Richtung einer möglichen Optimierbarkeit.

It might be argued that the drift in mass culture toward the southeast [Anm. im oben skizzierten Quadrantensystem: "sozial undurchsichtig"] is not a particularly fit strategy, as the propensity for adaptation found in the north is lost. We might assume that because we've spent most of our evolutionary history in the north, the "best" behaviors, in terms of fitness, are the ones that become the most popular. It could be argued that important health decisions ought to lie in the northwest, and in traditional societies it seems that's the case – such decisions are not strongly socially influenced. (Bentley et al. 2014, S. 75)

Das neue Paradigma der Big Data Analyse hätte die Beschreibung sozialer Phänomene wie folgt verändert:

Now, in the big-data era, the process of science and invention – the creation of ideas – has become well-documented. If science proceeds ideally (Kitcher 1993), it ought to plot in the northeast, where social learning is well-informed and seminal works by academic leaders are selectively adopted and developed by followers (Rogers 1962). Indeed, bibliometric studies show the most successful (highly cited) scientific mentors are those who train fewer protégés (Malmgren et al. 2010), which confirms high J_t (more social interaction) together with high b_t (more transparency). Also consistent with characteristics of the northeast, citations to scientific articles, and also patents, exhibit a highly right-skewed distribution with only slow turnover in the ranked popularity of citation rates (Bentley & Maschner 2000; de Sola Price 1965; Stringer et al. 2010). (Bentley et al. 2014, S. 70)

Zur Erinnerung, der Quadrant im Nordosten beschreibt inherent transparente, sozial beeinflusste Entscheidungen. Es mache, so Bentley - oftmals Sinn dahingehend zu optimieren, solange in gesellschaftlichen Entscheidungsprozessen zumindest einige wenige Individualentscheidungen weiterhin unabhängig davon getroffen werden.

For example, fishermen who always copy other, perhaps more successful fishermen can get stuck in a poorer part of a fishery and fail to locate better areas (Allen & McGlade 1986). Efficient communal fishing requires some individual boats to randomly probe other areas of a fishery than the ones that look apparently the best based on past catch experience. (Bentley et al. 2014, S. 68)

Es sei ebenfalls angemerkt, dass es sich hierbei um einen Versuch handelt, klassische Datenanalyse (durch einfache algorithmische Auswertung abdeckbar) mit aktuellen Erkenntnissen aus den Verhaltenswissenschaften, insbesondere der Verhaltensökonomie zu verbinden, um die Interpretationsfähigkeit von auf reiner Korrelation beruhenden, aus der Big Data Analyse gewonnenen Erkenntnissen zu erhöhen, oder überhaupt erst herzustellen. Bentley et al. nehmen zu diesem Zweck auch auf die Arbeit von Kahneman Bezug, der hier, als Vertreter einer neueren Entscheidungstheorie, später noch Erwähnung finden wird.

Die Ambitionen des genannten Ansatzes fallen durchaus breit aus und würden in ihrer Aussagewirkung Gesellschaftsprozesse berühren.

The map provides a means for evaluating population-level trends in the kinds of decision-making categories noted earlier – voting, opinions on climate change, mating, health care, consumer trends – which we see beginning to move to the south-east. [Anm.: sozial intransparent]

This can be important in a world where policy may not match the way human decisions are made. Much of traditional social science and policy has used the northwest as its base assumption that public behavior is dictated by cost/benefit ratios and incentives, even if slightly flawed or biased in perception (e.g., Kahneman 2003). This comprehensive practicality is why the map is not just another theoretical classification scheme. It allows a data-driven study question, reframed as a hypothesis, to be tested with appropriate datasets by examining whether bt, the degree of transparency of payoffs, decreases through time and whether Jt, the intensity of social influence, increases through time[...] (Bentley et al. 2014, S.74)

Bentley et al. reflektieren an dieser Stelle auf die Wirksamkeit von politischen Richtlinien um ein gewünschtes Gesellschaftsverhalten zu erreichen und umschreiben konkrete Bereiche in denen das Entscheidungsverhalten von Gruppen, erfahrbar gemacht durch die Big Data Analyse, nicht mehr durch stark rationale Kriterien beeinflusst seien. Eine Re-Evaluation der politischen Strategie wird nahegelegt.

Die Journal-Veröffentlichung des Artikels von Bentley et al. ging mit einem „call for papers“ einher, sodass ein wertvoller Blick auf die direkten Reaktionen innerhalb mehrerer wissenschaftlicher Richtungen auf die Veröffentlichung, geworfen werden kann.

Die erstgereichte Replik, nach der redaktionellen Bearbeitung des Journals, stammt von fünf Mitgliedern des Berliner Max-Planck-Instituts für Bildungsforschung, darunter auch, in hinterer Reihung, Gerd Gigerenzer, dessen Forschungsgebiet ebenfalls begrenzter Rationalität und Heuristiken in Entscheidungsprozessen beinhaltet.

Sie beginnt wie folgt -

We applaud Bentley et al. for postulating a map of the environment in the form of a two-by-two classification. Too much of psychological theorizing still focuses on internal factors alone, such as traits, preferences, and mental systems – a kind of theorizing that social psychologists once labeled the “fundamental attribution error.” Yet Bentley et al. are in danger of committing the opposite error: to theorize without regard of the cognitive processes. (Analytis et al. 2014, S. 76f)

und ist geeignet als Blaupause für den weiteren Kanon (soweit ein solcher attestiert werden kann) der Repliken hinzuhalten.

Die Veröffentlichung weist darauf hin, dass abhängig von einem induzierten Bias im Entscheidungsprozess einer klar innerhalb von in dem System nach Bentley et al. zu verortenden Anordnung (10.000 simulierte Proponenten), Effekte erzielt werden konnten, die die beobachtete Population unabhängig ihres eigentlichen Entscheidungsumfeldes (Entscheidungstransparenz zB. vorgegeben) die Charakteristik eines anderen Quadranten annehmen lies – und hält somit eine Erweiterung um ein Kriterienschema zum Prozessverlauf von Entscheidungsfindung für angebracht.

Die implizit genannte Gefahr darin den Menschen nicht mehr als frei denkendes Individuum zu begreifen, ist ebenfalls ablesbar.

Weitere Repliken attestieren mögliche Looping Effekte die, selbstverstärkend durch ihre Adaption in der Population erwartetes Entscheidungsverhalten beeinflussen, oder weisen darauf hin, dass Verhaltensänderungen, beispielsweise durch sich plötzlich ihrer Entscheidungsmuster bewusst gewordenen Populationen indiziert werden können. Beziehen sich also auf mögliche Störfaktoren, die innerhalb dieses Kriterienrasters nicht beachtet wurden.

Als weitere Kriterien für eine mögliche z-Achse, ein in den Reaktionen spürbar populärer Ansatz, werden unter anderem Emotion, oder die Netzwerkstruktur (Social Graph) der einzelnen Datensubjekte vorgeschlagen.

Gesamt fällt die Reaktion auf das vorgeschlagene Auswertungsschema positiv aus, insbesondere, dass viele Datasets von dieser Form der Betrachtung profitieren würden, wird anerkannt.

In der Gesamtbetrachtung bleibt zu vermerken, dass alle Repliken die darauf abzielen das Entscheidungsverhalten von Gruppen stärker miteinzubeziehen um mögliche Fehlerquellen auszuschließen, ein starkes Bedürfnis zeigen einen Kausalzusammenhang zwischen dem tatsächlichen Entscheidungsverhalten und der ablesbaren Verteilung zu etablieren, der ihrem Bild von Entscheidungsfindung entspricht.

Die deutlichste, jedoch nicht minder eindrucksvolle Kritik des Ansatzes von Bentley et al. stammt aus Wien.

Fred L. Bookstein, Anthropologe und Statistiker an den Universitäten Wien und Washington, erinnert an einen Ausspruch Korzybskis und verweist darauf, dass „die Karte niemals das Territorium sei“. Anschließend zeigt er mögliche Mängel in einzelnen theoretischen Überlegungen des Vorschlages auf. Die definierten Achsenrichtungen seien eine einfache Übertragung eines theoretischen Konzeptes von Levi-Strauss das zwei beliebige verbale Gegensätze (unabhängig und sozial, sowie transparent und opaque) so gegenüberstelle als seien sie absolute Gegensätze „[as] fundamental as water versus land, male versus female, light versus darkness, alive versus dead, or raw versus cooked, then converted into abstract propositions.“ (vgl. Bookstein 2014, S. 78)

Die relative Vergleichbarkeit der Gegensatzpaare in ihrer Relation sei nicht gegeben, das Modell impliziere einen Zustand bei dem die Optimalorientierung im jeweiligen Sektor an seinem Stamm (seiner Mitte) auszurichten sei. Bewegungen innerhalb der Quadranten seien zur Freude der Theoretiker nur parallel zu den Achsen möglich,

eine Mitte zwischen den Quadranten oder eine Möglichkeit eine Bewegung dahingehend zu interpretieren existiere erst gar nicht. Die als Testbeispiele von Bentley et. Al verwendeten Datasets, die beide zeitliche Verläufe beschreiben, könnten in komplett anderen geometrischen Figuren dargestellt und ebenfalls in vier Quadranten verortet werden, sodass ihre visuelle Repräsentation komplett andere Interpretationsmuster und Analyseansätze nahelege. Dies bricht einen der Hauptansätze von Bentley et al. die in der schnellen visuellen Repräsentation und der darauf basierten Zuweisbarkeit zur weiteren forschungsrelevanten Behandlung einen der wesentlichen Aspekte ihres Modelles sehen.

Bookstein sieht in der Proklamation wie Datasets zukünftig gefiltert werden sollen eine unzulässige Missrepräsentation der sozialen Realität, ohne Rücksicht auf daraus erwachsende Risiken.

For a social world in which the idealization of types need not correspond to any idealization either of their relationships or their representations in the minds of the agent, the imposition of a false symmetry of this kind entails the embrace of a potentially fatal rhetorical risk without any concomitant benefit. (Bookstein 2014, S. 79)

6. Eigenheiten in den Limitierungen von Big Data

Wie aus dem soeben skizzierten Beispiel einer scharfen Debatte innerhalb der Sozialpsychologie abzulesen, zeichnet sich der Versuch einer zukünftige Verbindung von Datenwissenschaftlern, Statistikern und Sozialwissenschaftlern ab, um gemeinsame Theorien zu entwickeln, die es erlauben Zusammenhänge in Datasets auch behavioristisch zu hinterfragen und dadurch besser fassen zu können.

Der argumentierbar sichtbarste Grund, warum Big Data Proponenten an dieser interdisziplinären Verbindung interessiert sind, ist darin zu suchen über Korrelationsaussagen durch beobachtete Häufungen und Muster in Verhaltensprozessen hinauszukommen.

Es existiert zwar die Argumentationslinie innerhalb der Gruppe der Befürworter, dass ab einer sehr hohen Korrelationsdichte ein Kausalitätsnachweis hinfällig wäre – aber diese Vorstellung ist selbst im interdisziplinären Diskurs umstritten.

So bezeichnet Kate Crawford - leitende Wissenschaftlerin des Microsoft Research Centers in der Forschungsgruppe „Social Media Collective“ - die eben genannte Position als Datenfundamentalismus - Massendaten seien nicht objektiv, sie seien ein Kind menschlichen Designs. (vgl. Crawford, 2013) Wenngleich diese Auslegung nicht alle Positionierungen innerhalb des wissenschaftlichen Diskurses widerspiegelt. So sprechen beispielweise Cukier (vgl. Cukier et al., 2013) und Mayer-Schönberger (vgl. Mayer-Schönberger, 2013) davon, dass aus Korrelationen gewonnene Erkenntnisse bereits zu einer Anwendung geführt werden sollten, bevor ein kausales Erklärungsmuster für den Zusammenhang besteht. Der Zusammenhang wäre a priori von Bedeutung, erst an zweiter Stelle käme die versuchte Erklärung.

Crawford erläutert ihren Standpunkt anhand einiger Beispiele in denen Big Data Ansätze zur Voraussage von Entwicklungen zumindest operativ in ihren Ambitionen gescheitert sind, und wie dies durch angelegte Defizits in der Erfassung der Datasets selbst zu erklären sei.

Eines der von ihr dazu herangezogenen ist die vielerseits beachtete Entwicklung von Google Flu Trends.

6.1 Google Flu Trends

Google Flu Trends startete 2008 als eines der experimentellen Webservices des Unternehmens und verfolgte die Zielsetzung das Auftreten von Grippe-Epidemien in mehreren Ländern der Erde aufgrund einer Korrelation zwischen Arztbesuchen und Suchanfragen innerhalb einzelner Regionen vorauszusagen.

Im Zeitraum 2009/2010 gelang dies eindrucklich, als das Unternehmen korrekte Voraussagen über den regionalen Ausbruch von Grippewellen in den US, beinahe zwei Wochen vor den Prognosen des Centers of Disease Control and Prevention (CDC) veröffentlichen konnte.

In den Jahren 2011 bis 2013 lies die Genauigkeit der Google Prognose jedoch deutlich nach, und produzierte Abweichungen zu den CDC Statistiken in einem Schwankungsbereich von beinahe hundert Prozent. Das Unternehmen selbst äußerte sich nicht im Detail zu den dafür verantwortlichen Gründen, es ist jedoch bekannt, dass der Algorithmus vor dem Hintergrund eines Musters von Datasets aus der Periode zwischen 2003 und 2007, unter anderem basierend auf 45 spezifischen Suchbegriffen der Suchmaschinennutzer, arbeitete und es wird angenommen, dass sich seit dieser Periode die Häufigkeit der Suchanfragen, ausgelöst durch Medienreportagen in der Grippesaison, unerwartet verändert haben dürfte. (vgl. Lazer et al. 2013, S. 127f).

Versuche für diesen nicht bestätigten Faktor zu korrigieren endeten letztlich in einem Misserfolg. (vgl. ebd.)

6.2 Aus dem Kontext

Crawford und Boyd führen in einem Artikel von 2012¹⁹ aus, wie Big Data große Teile seiner Bedeutung verliert, wenn Daten nicht mehr in einem detaillierteren Kontext betrachtet werden können. So seien in den vergangenen Jahren hunderte Forscher in den Social Media Bereich geströmt, um Verbindungen zwischen Individuen zu analysieren und seien diesbezüglich beteiligt daran gewesen, dass der Begriff „Social Graph“ auch innerhalb der Industrie einen Boom erfuhr – jedoch unterscheide sich der Social Graph im Detail sehr stark von dem was Soziologen und Geisteswissenschaftler seit den 30er Jahren in der Form von Soziogrammen und „sozialen Verwandtschaftsnetzwerken“ beschrieben. Die persönlichen Netzwerke hätten eine andere Qualität.

Crawford und Boyd beschreiben zwei neue Ausprägungen von interpersonellen Beziehungen in sozialen Netzwerken, die sie als artikulierte Netzwerke und Verhaltensnetzwerke beschreiben. Artikulierte Netzwerke bestünden aus vornehmlich behaupteten Beziehungen aufgrund technischer Bedingungen. Email Adressbücher, Telefonbücher, eine Mehrzahl an Freunden in Sozialen Netzwerken und dergleichen mehr. Die Beziehungen sind technisch vorhanden, werden jedoch nicht gepflegt. Unter Verhaltensnetzwerken verstehen Crawford und Boyd Netzwerkbeziehungen die durch Interaktionen belegt sind. Kommunikationsmuster, Bewegungsmuster, das gemeinsame „getagt“ werden auf Photos. Diese seien zwar für viele Betrachtungsansätze sehr nützlich, würden qualitativ jedoch kaum eine Aussage über die Stärke einer dieser Beziehungen erlauben. Beziehungsstärke, beispielsweise durch Kontaktfrequenz abbilden zu wollen, sei nicht ratsam – da Individuen andere Kriterien heranziehen, um die Stärke einer wechselseitigen Beziehung zu bewerten.

Datensets seien demnach nicht generisch. (Crawford versteht sie als a priori „designed“.) Es mache Sinn Abstraktionen zu diskutieren, jedoch nur, wenn auch der jeweilige Kontext mit bedacht wird.

¹⁹ Crawford & Boyd (2012) Critical Questions for Big Data

Kontext sei in der Betrachtung von Massendaten sehr schwer herzustellen, und ginge sehr schnell verloren, wenn Daten durch Selektion reduziert werden um in ein Modell zu passen. (vgl. Crawford & Boyd, 2012)

7. Der unsichtbare Algorithmus

Vor dem Hintergrund eines Eingriffs der aus dem spezifischeren Kontext heraus problematischer scheint, als in einer reinen datenorientierten Logik, kann auch Mayer-Schönbergers Kommentar zur Abwendung von Finanzmarktkrisen gesehen werden.

Zuvor sei relativierend angemerkt, dass Mayer-Schönberger in einer Mai Ausgabe von Die Presse, 2009 mit folgendem Ausspruch zitiert wird –

Und angesichts der Finanzkrise beunruhige ihn [Anm.: Mayer-Schönberger] die grassierende Meinung in der Bevölkerung, „dass der Markt versagt hat. Planwirtschaft ist sicher nicht die Lösung.“ (Kordik 2009)

Dennoch handelt es sich bei jedem Bemühen einem Börsencrash bereits im Vorfeld zu verhindern, und bestehe diese auch nur in einem automatisierten Eingriff nach der Erkennung eines bestimmten Musters, um eine nicht zu unterschätzende steuerungspolitische Maßnahme.

Dass Reaktionen auf solche Eingriffe ebenfalls nicht unbedingt gemäßigt ausfallen, verdeutlicht an dieser Stelle ein Auszug aus einem Essay im kulturevolutionären Adbusters Magazine.

In 2010, when the Dow Jones Industrial dropped 1000 points in under a minute, the biggest one-day point decline in history, it received far less attention than it deserved, because everything returned to normal a few seconds later. Now, miniature flash crashes occur constantly throughout the day. But this crash was a turning point, demonstrating that something had changed. That something was that the

neoliberals had achieved what communists, socialists and Christians never could: they made their god real, and in doing so, achieved their utopia. They just didn't let the rest of us in on it.

The critical flaw in Hayek's vision of the [invisible] hand was that a "central body" could never gather enough information. (Haddow, 2015)

Haddow, ein freier Journalist des Guardian, führt im Artikel auch einen kurzen Konnex zu der von Hayek mitbegründeten Mont Pelerin Society zu deren erstem Treffen unter anderem Milton Friedman, Wilhelm Röpke, und Karl Popper geladen waren.

Unabhängig von den implizierten Wertungen innerhalb der zitierten Betrachtung die hier, zur Wahrung des Prinzips der Meinungsvielfalt, als Beispiel für eine Position des gegenläufigen politischen Feldes erwähnt wird, findet sich darin jedoch auch eine historische Zuweisung, wie die Sozialwissenschaft in der Etablierung eines neuen Gesellschaftsverständnisses mit eingebunden ist.

Wann immer sich neue handlungsleitende Prinzipien in der Gesellschaftsentwicklung ausbilden, sind die Sozialwissenschaften gefordert diese zu hinterfragen und wissenschaftstheoretisch zu untermauern, um den Mechanismus dahinter transparent und gesellschaftlich verständlich werden zu lassen.

Es stellt sich nun direkt die Frage, inwieweit die sehr dispersen Positionen von Big Data Befürwortern und Kritikern wissenschaftstheoretisch innerhalb der Sozialwissenschaften verortet werden können.

Mit Popper eröffnet sich hier nachfolgend das Feld der Wissenschaftstheorie.

8. Wissenschaftstheorie

8.1 Korrelation contra Kausalität

In einem 2013 erschienenen Essay in dem vom Council on Foreign Relations publizierten Magazin „Foreign Affairs“ beschreiben Cukier et al. das Verhältnis von Big Data Anwendungen zur Korrelation, gegenüber der bisherigen wissenschaftstheoretischen Bedeutung von Kausalität wie folgt.

[The] two shifts in how we think about data — from some to all and from clean to messy — give rise to a third change: from causation to correlation. This represents a move away from always trying to understand the deeper reasons behind how the world works to simply learning about an association among phenomena and using that, to get things done. (Cukier et al., 2013)

Dieser sehr pragmatisch formulierte Ansatz beschreibt, so wir seiner Gültigkeitsbehauptung folgen, einen der größeren Brüche von Wissenschaftstheorie in der Wissenschaftsgeschichte der letzten Jahrzehnte.

Das Problem, so Cukier weiter, bestünde darin, dass kausale Zusammenhänge sehr schwer zu erfahren seien und sie sich oftmals, nachdem sie die Wissenschaft identifiziert habe, als nicht viel mehr als „selbst-beglückwünschende Illusionen“ erweisen würden. Cukier sieht das durch die Forschungsrichtung der Behavioral economics bestätigt.

Behavioral economics has shown, that humans are conditioned to see causes, even where none exist. So we need to be particularly on guard to prevent our cognitive biases from deluting us; sometimes, we just have to let the data speak. (Cukier et al., 2013)

Cukiers Auffassung geht einher mit einem Realitätsverständnis das nicht mehr durch die Notwendigkeit von Samples in der empirischen Betrachtung geprägt ist, sondern annimmt, dass durch die reine Menge an Informationen, Theorien über allgemeine Zusammenhänge nicht erst aus extrapolierten Gesetzmäßigkeiten abgeleitet werden müssen, sondern a priori für die betrachteten Populationen, die in ihrem Umfang weit größer sein können als das was den Sozialwissenschaften bisher durch konventionelles statistisches Sampling erfahrbar gemacht wurde, erst einmal bestehen würden.

Daran knüpft sich auch ein bestimmtes Theorieverständnis von Wiederholbarkeit, wonach der Suche nach wiederkehrenden Gesetzmäßigkeiten keine besondere Notwendigkeit mehr zukommt, da der Grund weshalb ein Ereignis eintritt nicht mehr von ausschließlicher Bedeutung ist, wenn gleichzeitig angenommen oder sogar prognostiziert werden könne, dass es eintritt. Unabhängig von seinem Charakter da die Aussage bezüglich des Auftretens auch noch Bestand habe, wenn sich diese, beispielsweise abhängig von einem zeitlichen Faktor ändere. Das Ziel bestehe somit nicht mehr in einer Erklärung von oftmals wandelbaren Phänomenen, sondern vielmehr im Aufzeigen der selbstigen. (vgl. Cukier et al., 2013)

Um diesem Theorieverständnis folgen zu können, ist es notwendig Abfrageintervalle so weit als möglich zu verkürzen, was letztlich in einem größeren Datenbestand mündet. Solche Versuche finden, so Cukier, bereits heute ihre Anwendung.

Cukier nennt ein Beispiel von kanadischen Forschern die einen Big Data Ansatz entwickeln, um eine Früherkennung von Infektionen bei frühgeborenen Babys zu ermöglichen. Diese messen nicht nur 16 Vitalparameter bei jedem der Kinder, sondern tun dies fortwährend, sodass in dem Prozess 1000 Datenpunkte in der Sekunde anfallen. Die Forscher interessieren dabei vor allem Korrelation zwischen kleinsten Abweichungen von der Norm der Messergebnisse und später auftretenden Komplikationen. In Teilen sei dies bereits erfolgreich, und über eine zunehmende Dauer würde dieser Zugang es den behandelnden Ärzten eventuell sogar ermöglichen, durch mehrere aufgezeichnete Betrachtungen zu einer Erklärung für das warum, einer Kausalität eines Vorganges zu finden. Jedoch bestehe auch ohne eine sol-

che eine unbedingte Indikation zur Verwendung der Methode, da die Big Data Analyse in der Zwischenzeit „bereits Leben retten könne“. (vgl. Cukier et al., 2013)

Aus dieser Besonderheit, auf den Echtzeitcharakter einer Auswertung zu fokussieren zu können, ziehen einige Vertreter des Big Data Paradigmas, eine Begründung Korrelationszusammenhänge bereits als handlungsanleitend verstehen zu können. Das Verständnis eines Feldes ergibt sich dadurch nicht mehr aus einer Abfolge von Grund und Folgewirkung, sondern der Erhebung von möglichen, statistisch signifikanten, Zusammenhängen. Auch ohne dabei ein Ende der Bedeutung von Kausalität in der Forschung behaupten zu müssen, wird hier ein neuer Zugang beworben.

Es scheint unwahrscheinlich, dass dieser Ansatz ein Prinzip von geringer Bedeutung bleibt, denn es existieren konkrete Anwendungsbeispiele die ihn als Nutzungsvoraussetzung in sich tragen und dabei zumindest einen Teil eines sich ändernden Verständnisses im Rahmen von Technologieentwicklung repräsentieren.

Unter anderem im Feld „machine learning“ das eine treibende Kraft hinter Anwendungsansätzen wie speech to text, automatisierten Übersetzungen, oder den Wegfinderroutinen selbstfahrender Autos darstellt. Ein Verständnis von dem was korrelationsgetriebener Forschung, ohne einen Bedeutungszusammenhang zu benötigen, zu bewerkstelligen möglich ist, könnte einen Teil der öffentlichen Diskussion nachhaltig prägen.

Vor dem Hintergrund dass Korrelationsaussagen selbst in einer neuen Qualität zu einem forschungsleitenden Ansatz beim Verständnis von sozialen Phänomenen aufsteigen könnten, stellt sich nun unmittelbar die Frage, in wie weit dies mit dem gängigen Verständnis in der Wissenschaft vereinbar ist.

In den Sozialwissenschaften hat sich nach dem Positivismus Streit von 1961 eine gewisse Dialektik der Methoden ausgebildet in der eine Koexistenz von quantitativen und qualitativen Ansätzen möglich ist. Der handlungsleitende Rahmen wird da-

bei insbesondere von zwei Vertretern des Post-Positivismus bestimmt. (vgl. Herzog 2012, S. 52, 69f)

8.2 Das Wissenschaftsverständnis nach Karl Popper

Karl Popper gilt seit dem Erscheinen seiner „Logik der Forschung“ als Begründer einer falsifikationszentrischen Methodenlehre, die forschungsanleitende Hypothesen hinsichtlich einer möglichen Falsifikation prüft. Die Prüfung erfolgt dabei im Wissenschaftsalltag anhand intersubjektiver Kriterien was die Vergleichbarkeit von Erfahrungen in der Replikation von Untersuchungsdesigns sicherstellt.

Hypothesen können in diesem Paradigma nur innerhalb des kritischen Rationalismus falsifiziert werden, nämlich wenn sich die in der Folge gewonnenen Erfahrungswerte nicht mit den hypothetischen Vermutungen decken.

Lässt sich eine Hypothese nicht falsifizieren, erreicht sie theoretische Gültigkeit innerhalb der jeweiligen Forschungsrichtung – kann jedoch nicht für sich beanspruchen letztgültig wahrhaftig oder richtig zu sein, sie gilt nach dem jeweiligen Forschungsstand als nicht widerlegt und hat sich behauptet. (vgl. Herzog, 2012, 64f)

Diese Konzentration auf die Falsifikation von Erkenntnissen besitzt den unmittelbaren Vorteil, dass sie vollständig innerhalb der deduktiven Logik abhandelbar ist.

Vergleicht man unterschiedliche Positionen induktiver Logik, habe man es mit Realitätsbildern zu tun die miteinander konkurrieren, die jedoch nach Poppers Auffassung nicht aufgrund von Erfahrungswerten als wissenschaftlich valide bestätigt werden können, insbesondere da eine Allgemeingültigkeit schwer zu behaupten ist.

Popper beschreibt dies als Induktionsproblem - der Umstand, dass einer Hypothese oder Theorie aufgrund des Prozesses der Verallgemeinerung nicht mehr mit der selben Gültigkeit behauptet werden kann.

Hinsichtlich einer weiteren Kompatibilität mit dem Big Data Ansatz sei angemerkt, dass der sozialwissenschaftliche Erkenntnisansatz nach Popper nicht vom Zustandekommen einer Theorie oder Hypothese berührt wird.

Dies sei der Fall, da das Zustandekommen einer Theorie das Forschungsfeld nur in soweit berühre, als aus dem Gedanken in späterer Folge ein behauptetes Theorem wird, mit dem sich die Wissenschaft in Anlehnung einer Überprüfung der kausalen Gerichtetheit des Arguments, nachfolgend auseinandersetzen müsse.

Will man die Vorgänge bei der Auslösung des Einfalls nachkonstruieren, dann würden wir den Vorschlag ablehnen, darin die Aufgabe der Erkenntnislogik zu sehen. Wir glauben, daß diese Vorgänge nur empirisch-psychologisch untersucht werden können und mit Logik wenig zu tun haben. Anders, wenn der Vorgang der nachträglichen Prüfung eines Einfalls, durch die ja der Einfall erst als Entdeckung entdeckt, als Erkenntnis erkannt wird, rational nachkonstruiert werden soll: Sofern der Forscher seinen Einfall kritisch beurteilt, abändert oder verwirft, könnte man unsere methodologische Analyse auch als eine rationale Nachkonstruktion der betreffenden denkpsychologischen Vorgänge auffassen. (Popper, 1934, S. 6f)

Im Nebensatz trifft Popper jedoch möglicherweise eine Annahme die in dieser Form im Big Data Paradigma keine Gültigkeit mehr besitzt. Die Prüfung des erkenntnistheoretischen Einfalls stellt keine Vorbedingung für Entdeckung mehr dar, genausowenig wie für behauptete wissenschaftliche Erkenntnis.

In späterer Folge wird diese weiße Fläche innerhalb des Betrachtungsrahmens des Erkenntnisinteresses bei Popper, noch einmal deutlicher.

So hat wohl schon jeder Experimentalphysiker überraschende, unerklärliche „Effekte“ beobachtet, die sich vielleicht sogar einige Male reproduzieren ließen, um schließlich spurlos zu verschwinden; aber er spricht in solchen Fällen noch nicht von einer wissenschaftlichen Entdeckung (obwohl er sich vielleicht bemühen wird, Reproduktionsanordnungen für den Vorgang aufzufinden). Der wissenschaftlich belangvolle physikalische Effekt kann ja geradezu dadurch definiert werden, daß er

sich regelmäßig und von jedem reproduzieren läßt, der die Versuchsanordnung nach Vorschrift aufbaut. Kein ernster Physiker wird jene „okkulten Effekte“, zu deren Reproduktion er keine Anweisung geben kann, der wissenschaftlichen Öffentlichkeit als Entdeckung unterbreiten, denn nur zu bald würde man auf Grund des negativen Resultats der Nachprüfungen die „Entdeckung“ als ein Hirngespinnst ablehnen. (Popper 1934, S. 19)

Wenn Mayer-Schönberger vom Prinzip „N = all“ oder Proponenten der Big Data Analyse allgemein von Samples sprechen, die groß genug seien um den Korrelationschluss vor den Kausalzusammenhang zu stellen, tritt eine erwartete Reproduzierbarkeit zunehmend in den Hintergrund. Auch kann die Reproduzierbarkeit unter Umständen erst garnicht theoretisch angenommen werden, da es dem Wissenschaftler selbst, selbst unter den besten Vorbedingungen, nicht möglich ist, eine eigene Erhebung in der selben Größenordnung anzustoßen, oder Zugang zu einem äquivalenten Datensatz aus anderer Quelle zu kommen – dadurch relativiert sich der Zwang zum unabhängig erbrachten Nachweis. Die Größe des Datensatzes ersetzt hierbei proklamativ, zumindest teilweise, das Bedürfnis zur unabhängigen Reproduzierbarkeit eines Ergebnisses.

Zudem existiert durch die Möglichkeit der immer weiter spezifizierbaren Detailtiefe einer Anfrage, ein neuer Ansatz, keine unbedingte Aussage über ein Sample mehr treffen zu wollen, sondern zu stark zur individualisierten Betrachtungen überzugehen (vgl. Mayer-Schönbergers Ansatz zum Abbau „gruppenbezogener Vorurteile“) der keine Behauptung von Allgemeingültigkeit mehr zugrunde liegt. Verstärkt wird dies durch den Ansatz einer Betonung der Aktualität von Korrelationen. Veränderungen sind dabei bereits implizit, ohne dass man es dem Forscher anlasten würde.

Es sei jedoch weiters festgehalten, dass es sich beim Ausspruch Poppers nicht um eine Behauptung handelt, was Wissenschaft zu sein habe, sondern er eine angenommene Grundbedingung darstellt, die im popperschen Verständnis in jedem Falle erfüllt sei. Das oben genannte manifestiert sich also nicht in einer Kritik an der Wissenschaftstheorie Poppers, sondern lediglich unter dem Wegfall einer angenomme-

nen Grundbedingung einer seiner Erläuterungen (siehe: „der belangvolle physikalische Effekt“).

Popper definiert in „Logik der Forschung“ weiters aus, dass ein Streit darüber ob es „nicht wiederholbare, einzigartige Vorgänge“ gäbe, innerhalb einer Wissenschaft nicht lösbar wäre. Dieser Streit selbst sei „metaphysisch“. (vgl. Popper, 1934, S. 20)

Die Annahme dass dieser Ansatz die Wissenschaft nie berühren würde, kann vor dem aktuellen Hintergrund von Big Data Anwendungen als widerlegt gelten. Gerade die Frage in wie weit die aufgezeichnete Detailtiefe rückblickend Aussagen über nicht reproduzierbare Ereignisse zulasse, ist ein Kern der aktuellen Vereinbarkeitsdebatte.

8.3 Das Wissenschaftsverständnis nach Kuhn

Thomas S. Kuhn gilt gleichauf mit Popper als der zweite prominente Vertreter der Wissenschaftstheorie im Post-Positivismus. Kuhn prägte den Begriff des wissenschaftlichen Paradigmenwechsels der es bis in das alltägliche Sprachgut schaffte und heute untrennbar mit dem öffentlichen Verständnis von Wissenschaft verbunden ist. Sein 1962 erschienenes Hauptwerk „The Structure of Scientific Revolutions“ wurde zu einem der meistzitierten Werke der Wissenschaftsgeschichte und zu einem der einflussreichsten Bücher des 20. Jahrhunderts.²⁰

In „The Structure of Scientific Revolutions“ geht Kuhn im Gegensatz zu Popper davon aus, dass die Falsifikation einer These, im Rahmen des normalen wissenschaftlichen Betriebs (im Gegensatz zu dem in der Krise), den er als „puzzle solving“ beschreibt, nicht sofort durch auftretende Anomalien geschieht, sondern diese häufig ignoriert oder sogar wegargumentiert werden können, ohne dass dies zwangsläufig zu wissenschaftlichem Fortschritt, oder dem proklamierten Paradigmenwechsel

²⁰ vgl. Naughton (2012) <http://www.theguardian.com/science/2012/aug/19/thomas-kuhn-structure-scientific-revolutions> [Letzter Aufruf: 19. 04. 2015]

führt. Erst wenn sich diese Anomalien, die nicht aus Falsifikationen entstanden sein müssen, sondern beispielsweise auch durch neue, nicht mit alten Mustern vereinbare Erkenntnisse ausgelöst werden können („puzzle solving“ erweist sich als nicht erfolgreich), nicht im Rahmen des geltenden wissenschaftlichen Diskurses wegarargumentieren lassen, führe dies innerhalb eines bestehenden Paradigmas zu einer Krise. Kuhn beschreibt die Krise als einen Zeitpunkt außerordentlich aktiver Forschungstätigkeit und Meinungsvielfalt, als eine Periode in der alles versucht würde um diese letztlich zu überwinden. (vgl. Kuhn 1970, S. 60ff) Ausgehend von einer Krise kann es dazu kommen, dass vermehrt Missfallen gegenüber theoretischen Konzepten geäußert wird und Grundannahmen neu verhandelt werden. Aus diesem offenen Konflikt heraus entstünden neue Ideen, Methoden und Theorien, die erst dann zu einer Revolution bezüglich des alten Paradigmas führen können.

Dabei stehe es Wissenschaftlern frei sich zwischen unterschiedlichen Theorien zu entscheiden, wobei der Prozess dieser Entscheidung selbst nicht zwangsläufig rational ausfalle, sondern Zugehörigkeitsbekundungen oftmals mit Diskussionen einhergingen und nicht immer aus unabhängigen Betrachtungen erwachsen.

Kuhn selbst konkretisiert in einem späteren Essay mit dem Titel „Objectivity, Value Judgment, and Theory Choice“ (1977), nach wiederholt auftretender Kritik, dass der Prozess der Theorieübernahme von ihm als vor allem durch soziale Zwänge beeinflusst gesehen werden würde, dass dies nicht der Fall sei. Er selbst begreife diesen Prozess in dem sich die Wissenschaft als Gruppe neu orientiert, durchaus als informiert. Kuhn zieht für die Entscheidung zwischen zwei Theorien in weiterer Folge fünf Qualitätskriterien heran.

- Exaktheit, wonach die neuen Theorien aus neuen Forschungsergebnissen und Betrachtungen hervorgegangen sein müssen.
- Konsistenz, wonach die Theorien sowohl innerhalb ihrer inneren Logik funktionsfähig sein müssen, als auch in der Anknüpfung an andere, bestehende Erkenntnisse und Beobachtungen.

- Eine neue Theorie sollte eine gewisse Weitsicht ermöglichen, nicht nur ein sehr eng gefasstes theoretisches Problem zufriedenstellend lösen, sondern mehr Gestaltungsspielraum beanspruchen können.
- Sie sollte simpel sein, und verschiedene bisher isoliert stehende Einzelaspekte ordnen können.
- Sie sollte letztlich offen und sinnstiftend gegenüber neuen Theorien und Erkenntnissen im Feld sein, sowie dabei helfen neue Zusammenhänge aufzeigen.

Gesamtheitlich betrachtet, fällt nach Kuhn die Entscheidung zwischen zwei Theorien die unterschiedliche Paradigmen repräsentieren jedoch nicht strenggenommen objektiv aus. Die Gewichtungen der oben genannten Aspekte treffen einzelne Wissenschaftler selbst, und selbst bei exakt gleicher Gewichtung der Selektionsfaktoren, könne dies zu unterschiedlichen Entscheidungen von Wissenschaftlern führen. (vgl. Kuhn 1977, S. 357ff)

Kuhns Begriff vom Selbstverständnis einer Wissenschaft ergibt sich somit aus einem Übereinkommen von Meinungen und Positionen. Es ist darin unmöglich aus Leitsätzen objektivierbare „Wahrheiten“ abzuleiten oder auch nur als Zielvorstellung zu definieren. Es existiere auch kein letztgültiges Ziel das eine Wissenschaft anstreben würde, viel mehr gleiche der Entwicklungsprozess dem einer Evolution.

Mit dem Wechsel eines Paradigmas gehe nicht nur eine „neue Sicht auf die Dinge“ einher, die Wissenschaft arbeite in einer „völlig neuen Welt“. Dies sei dadurch gekennzeichnet, dass man neue Begrifflichkeiten forme, alten Begriffen teilweise eine neue Bedeutung zuspreche und dadurch die Vereinbarkeit der beiden Positionen selbst theoretisch verunmögliche. (vgl. Kuhn 1970, S. 112) Kuhn prägt hier den Begriff der Inkommensurabilität der die verunmöglichte Vergleichbarkeit aus der Sicht einer der beiden Perspektiven beschreibt.

Kuhn entwickelt sein Wissenschaftsverständnis aus einer wissenschaftshistorischen Grundbetrachtung und argumentiert vornehmlich mit historischen Beispielen, die nahelegen dass Wissenschaft im Allgemeinen keine homogenen, stetigen Entwicklungen vollziehe, sondern sich wissenschaftlicher Fortschritt mitunter sprunghaft manifestiere, auch unabhängig der jeweiligen zeitgeschichtlichen Einflüsse.

In einem weiteren Essay von 1974, „Second Thoughts on Paradigms“, beschäftigt sich Kuhn intensiver mit der Art und Weise wie uns Sinneseindrücke (die Kuhn explizit als „data“ benennt) zu Eigenschaftsattributionen bewegen. Kuhn beschreibt das Lernen von Attributionen am Beispiel eines Kleinkindes das mit seinem Vater das erste Mal einen zoologischen Garten besucht. Besonders interessant an der Stelle ist, dass sich Kuhn bereits Gedanken zu computational modeling und machine learning (mattern matching) macht und sogar selbst angibt an einer ersten Implementierung zu arbeiten.

By the end of the walk, features like the length and curvature of the swan's neck have been highlighted and others have been suppressed so that swan data match each other and differ from goose and duck data as they had not before. Birds that had previously all looked alike (and also different) are now grouped in discrete clusters in perceptual space.

A process of this sort can readily be modeled on a computer; I am in the early stages of such an experiment myself. A stimulus, in the form of a string of n ordered digits, is fed to the machine. There it is transformed to a datum by the application of a preselected transformation to each of the n digits, a different transformation being applied to each position in the string. (Kuhn 1974, S. 298)

Kuhn beschreibt in weiterer Folge wie das Kind, das von seinem Vater Attributionen präsentiert bekommt und dadurch lernt dessen Kategoriensystem von Gans auf sein Bild von der Wirklichkeit anzuwenden.

Das Gelernte für das Kind bestehe darin, symbolhafte Labels ausgehend von der Attribution des Vaters, bestimmten Mustern zuordnen zu können. Dabei habe das Kind selbst kein Verständnis dafür erhalten welchen vordefinierten Kriterien (was definiert eine Gans) das Label folgt, könne dieses Wissen daher auch nicht generalisieren, oder daraus beispielsweise ableiten wie ein anderes Label lautet (zB. das ist ein Schwan).

Das Kind, wie auch das Computer Modell lernen in diesem Fall anhand dessen was Kuhn „Exemplar“ nennt.

Kuhn stellt in weiterer Folge die bereits bei Popper aufgeworfene Kernfrage für das Verständnis der theoretischen Kompatibilität des Big Data Paradigmas.

Suppose scientists to assimilate and store knowledge in shared examples, need the philosopher concern himself with the process? (Kuhn 1974, S. 299)

Kuhn beantwortet die Frage wie folgt.

Der Philosoph habe die Möglichkeit Regeln durch konkrete Beispiele („Exemplars“) zu ersetzen und zumindest in der Theorie kann er erwarten, dass dies erfolgreich verläuft. Im Prozess selbst jedoch, wird er die Form des Wissens der Gruppe ändern. Was er tatsächlich vollziehe, sei jedoch eine Form der Verarbeitung von Sinneseindrücken durch eine andere zu ersetzen. Ist er dabei nicht äußerst achtsam, schwächt er dadurch die Wahrnehmung der Gruppe. Und selbst wenn er sehr achtsam sei, so Kuhn, wird er einige der zukünftigen Reaktionen der Gruppe auf Stimuli geändert haben.

Diese Ausführungen erinnern frappant an den bereits an anderer Stelle gefallenen Ausspruch Korzybskis „Die Karte ist nicht das Territorium“ – nur dass an dieser Stelle das Beispiel - und die durch das selbige repräsentierte Kausalentsprechung (Regel) gegenübergestellt werden.

8.4 Betrachtung der theoretischen Kompatibilität

Beide post-positivistischen Wissenschaftstheorien zeigen eine theoretische Kompatibilität zum Ansatz der Big Data Analyse. Postulierte Zusammenhänge wären nach Popper immer noch durch deduktive Logik falsifizierbar, auch ein postulierter Kausalschluss, so er gezogen wird, bleibt innerhalb eines post-positivistischen Verständnisses in ähnlicher Art und Weise widerlegbar. Hält beispielsweise die Datenbasis nicht, wären unhaltbare Theorien gegebenenfalls sogar noch schneller widerlegt.

Die Kernproblematik zeigt sich mehr im Detail, denn wenn Popper versucht mit Hilfe deduktiver Logik die subjektive Komponente von Theorien zu begrenzen und weniger stark einem individuellen Einfluss auszusetzen, befördert die Big Data Analyse – so sie einen Kausalzusammenhang herzustellen verlangt, beinahe explizit die Anwendung induktiver Logik. Inwieweit die daraus entstehende Zunahme einer Unschärfe im wissenschaftlichen Prozess durch die größere empirische Detaildichte eines Datasets unter Big Data abgedeckt ist, bleibt dabei noch ungeklärt. Vor dem Hintergrund dass Cukier (vgl. Cukier et al., 2013) in der menschlichen Natur einen Bias vorhanden sieht der das Treffen von Kausalschlüssen bevorzugt, und das in der Forschung der Behavioral Economics bestätigt sieht, verstärkt sich auch das Problem einer kausalitätsfreien Betrachtung von Forschungsergebnissen, da Wissenschaft an dieser Stelle nicht mehr handlungsleitend in Erscheinung tritt, und trotzdem nicht verhindert werden kann, dass vor dem Hinblick ihrer Ergebnisse Kausalschlüsse abgeleitet werden, die weiter von der Expertise entfernt sind, die der Forscher selbst aufgebaut hat.

Das zweite Detailproblem findet sich in der Zusammensetzung von empirischen Daten in beiden Verständnissen. Während der Forscher nach Popper immer dazu angehalten ist empirisch exakt zu arbeiten, sodass sich auch kleinere Fehler nicht auf die Repräsentativität des Ergebnisses auswirken, sieht sich der Big Data Forscher davon nur mäßig betroffen. Sofern eine Repräsentativität des Datasets angenommen werden kann, da beispielsweise das Sample groß genug ist um hinsichtlich „N = all“ zu divergieren, und zumindest grobe Versäumnisse beim Fassen eines Samples em-

pirisch ausgeschlossen werden können, betrachtet das Big Data Paradigma eine geringe Anzahl an falschen Datenpunkt als hinnehmbar.

Statisticians have shown that sampling precision improves most dramatically with randomness, not with increased sample size. In fact, though it may sound surprising, a randomly chosen sample of 1,100 individual observations on a binary question (yes or no, with roughly equal odds) is remarkably representative of the whole population. In 19 out of 20 cases it is within a 3 percent margin of error, regardless of whether the total population size is a hundred thousand or a hundred million. (Mayer-Schönberger 2013, Kapitel 2, fünfzehner Absatz)

Würde sich beispielsweise aus dem Zusammenziehen zweier Datasets eine versehentliche doppelte Einbeziehung eines Datums ergeben, wäre diese Unschärfe hinsichtlich der Anzahl der restlichen Datenpunkte für den Forscher vernachlässigbar. (vgl. ebd.)

Dadurch stellt sich die Frage nach der Vergleichbarkeit des jeweiligen Qualitätsverständnisses und welche Auswirkungen daraus in der Praxis entstehen.

Gesellschaftspolitisch betrachtet wäre eine verstärkte Akzeptanz von „false positives“ alleine durch die Wahl des jeweiligen Methodendesigns, jedenfalls nicht unproblematisch, denn im Gegenzug zu Cukiers frühgebohrenen Babies ist nicht jede Richtungsentscheidung dazu prädisponiert vor allem „Leben zu retten“.

Datasets unter dem Big Data Paradigma beherbergen noch ein weiteres Problem. Aufgrund der Größe und ihrer heterogenen Natur (zum Sammelzeitpunkt steht die Verwendung noch nicht notwendigerweise fest), müssen diese vor ihrer Analyse „gereinigt“ und in ein einheitliches Kodierungssystem überführt werden. Das passiert aktuell mehrheitlich durch die Datenwissenschaftler und Analysten selbst, die im Vorfeld dieses Prozesses durch erste Analyseanordnungen ein „Verständnis“ für die Korrelationen gewinnen. Zeigt sich zu diesem Zeitpunkt ein induktives Logikverständnis als handlungsleitend, besteht die Gefahr einer unerwünschten Einflussnahme – die dazu führen kann, dass Ergebnisse der Analyse nicht mehr valide blei-

ben. Insbesondere da die Erhebungskriterien nicht im Vorfeld festgelegt werden können, ist es vergleichsweise einfach möglich einen gesamten „Ast“ an Korrelationen auszusortieren, der sich später als wesentlich erwiesen hätte. In der Praxis sei dieses Problem nicht durch einen böswilligen Vorsatz geleitet, es sei schlichtweg für den Analysten sehr einfach möglich die falschen Entscheidungen zu treffen. (*vgl. Simmons et al. 2012*)

Unter Datenwissenschaftlern existieren aufgrund dessen Überlegungen die Freiheitsgrade von Analysten bewusst einzuschränken. (*vgl. edb.*)

Ein weiteres Detailproblem besteht, wie weiter oben ausgeführt darin, dass Poppers Auffassung davon, dass das Entstehen einer Theorie oder Hypothese nur in sofern für die Sozialwissenschaft von Bedeutung sei, als dass sie den daraus resultierenden proklamierten Kausalzusammenhang zu falsifizieren suche. Dies deckt sich nicht mit den momentanen Bestrebungen auf wissenschaftlicher Ebene. Ausgehend von einer breiten Datenbasis erwächst zunehmend das Interesse an induktiver Logik zur Erfassung von Bedeutungszusammenhängen, die sich gegebenenfalls nicht unabhängig falsifizieren lassen. Die Art und Weise der Theoriebildung muss vor diesem Hintergrund in einer neuen Weise reflektierbar gemacht werden, so sie die Sozialwissenschaft zu interpretieren sucht.

Das Wissenschaftsverständnis nach Kuhn legt eines mit Sicherheit nahe, dass es sich beim Big Data Ansatz um ein neues Wissenschafts-Paradigma handelt. Als solches erfüllt es auch die von Kuhn genannten Qualitätskriterien für eine divergierende Wissenschaftstheorie. So die Zusammensetzung des Datasets keine gravierenden Mängel aufweist, ist sie exakt, denn sie geht aus empirischen Daten hervor, es existiert eine Konsistenz der inneren Logik (zB. Korrelation erweise sich als bedeutsamer als Kausalität), Anknüpfungspunkte zu bestehenden Theorien sind vorhanden - das Big Data-Paradigma verbreitet sich gerade in den unterschiedlichsten Wissenschaftsrichtungen, der Gestaltungsraum ist nicht eingengt und offen für neue Forschungsansätze, sowie Erkenntnisse ist sie als Methode ebenso.

Das Wissenschaftsverständnis nach Kuhn sorgt in diesem Zusammenhang genauso wie das von Popper auch für mögliche Probleme, wenn damit eine konkrete durch die Big Data Analyse gestützte Theorie betrachtet wird, beispielsweise, da es der Wissenschaftstheorie die Grundlage für eine objektivierbare, zielorientierte Perspektive abspricht. So wären die erhobenen Datenbestände, in dem von Kate Crawford postulierten Verständnis - konstruiert - und die Gültigkeit der daraus getroffenen Leitsätze ebenfalls nur durch ein Übereinkommen, ein Einverständnis zwischen Wissenschaftlern gegeben.

Kuhn selbst belegt dies jedoch nicht mit einer negativen Konnotation, sondern bezeichnet es als wünschenswert –

“What better criterion could there be,” I asked rhetorically, “than the decision of the scientific group?” (Kuhn 1977, S. 359)

Zusammenfassend lässt sich festhalten, dass in der Detailbetrachtung Probleme entstehen, die jedoch im Rahmen eines aktiven Diskurses der Vertreter beider wissenschaftstheoretischer Richtungen vereinbar scheinen. Trotzdem zeigen sich Parallelen zum Positivismusstreit in den Sozialwissenschaften.

9. Urteilsheuristiken

9.1 Kahneman

Kahneman der von unterschiedlichen Vertretern des Big Data Paradigmas bereits mehrfach referenziert wurde, unterstreicht in der von ihm begründeten Forschungsrichtung der Behavioral Economics die Auffassung, dass das Treffen von Kausalschlüssen ein schwieriges Unterfangen sei, in dem er das Versagen von erfahrungsgeleiteter Intuition im Rahmen seiner wissenschaftlichen Tätigkeit anhand unterschiedlicher Experimentalanordnungen verdeutlicht. (vgl. Gradinarua 2014)

Kahneman verortet die Intuition im selben System wie die Expertenmeinung die aus einer Intuition heraus - geleitet durch Erfahrung - entsteht. Diese auf Heuristiken basierende Entscheidungsebene leite unser Handeln in einer Weise, die vielfach überraschend und wenig rational sei.

Die intuitive Ebene kann dabei als eine Art automatisierter Bias gesehen werden, der jedoch gleichzeitig auch Expertise (Einschätzungen aufgrund vergangener Muster) beinhaltet. Dieser Bias sei grundsätzlich aktiv und es benötige eine explizite Anstrengung, oder ein unerwartetes Element um ihn zu Gunsten einer rationalen Überlegung beiseite zu stellen.

Kahneman teilt, laut eigener Aussage, sein Basis-Verständnis von Intuition mit einer Definition von Herbert A. Simon²¹, wonach Intuition gleich Wiedererkennen sei – interessanter Weise beschreibt er das Einsetzen von Intuition in einer sehr ähnlichen Weise wie Kuhn das Prinzip der Attribution von Bedeutung mit dem Kind erklärt, dass die Zuweisung eines Begriffs anhand eines „Exemplars“ (in der Bedeutung von Kuhn) erlernt hat.

Valid intuitions develop, when experts have learned to recognize familiar elements in a new situation and to act in a manner that is appropriate to it. Good intui-

²¹ vgl. What Is an “Explanation” of Behavior?

tive judgments come to mind, with the same immediacy as "doggie". (Kahneman 2011a, S. 16)

Kahneman versteht darunter, wie Kuhn, eine erlernte Attribution die eintritt, sobald jemand die Regel zur Identifikation eines Musters verinnerlicht hat. Eine Expertenintuition bilde sich, so Kahneman, wenn jemand in längeren Kontakt mit einem bestimmten Feld komme.

Im Gegensatz zu Kuhn, finden wir bei Kahneman jedoch eine explizite Wertung, welcher erkenntnistheoretische Stellenwert einer derartigen Attribution, fortan als Intuition benannt, zuzuschreiben ist.

Expertenintuition leite sich nicht von echter Expertise ab und erweise sich immer wieder als nicht sehr zuverlässig. Werden auf ihrer Basis Entscheidungen getroffen setze primär ein Gefühl von subjektiver Zuversicht ein und nicht ein Bedürfnis kritischer Reflexion.

Dazu Kahneman selbst in einem Vortrag:

Subjective confidence is actually not a judgement at all it is a feeling. It is a feeling that people have. And I think I know where the origin of that feeling is, and it is – "System 1"²² if you will, assessing the fluency of its own processing. Assessing the coherence of the story it has created, to deal with the current situation. And if the story is coherent, confidence is high.

²² System 1 beschreibt im Rahmen eines von Kahneman geschaffenen theoretischen Modells, zudem er selbst anmerkt, dass es als Konzeptionshilfe gedacht ist und er nicht annimmt, dass es dafür eine konzeptuell ähnliche Entsprechung in unserer Physiognomie gibt – eine von zwei Entscheidungsebenen. System 1 sei eine schnelle (agiert immer zuerst), automatische (lässt sich nicht deaktivieren) und intuitive (ruft keine konzeptuellen Modelle ab) Entscheidungsebene mit deren Hilfe wir Heuristiken bilden, die von System 2 relativiert, überstimmt oder zumindest relativiert werden können. System 2 ist analytisch, vernunftbasiert und viel langsamer. In System 2 können wir erlernte Regeln oder Vergleichsmodelle zum Einsatz bringen das erfordere jedoch auch mentale Anstrengung (Pupillen weiten sich beispielsweise). System 2 kommt beim Treffen von Entscheidungen nicht zwangsläufig zum Einsatz – wie auch anhand des oberen Zitates abgelesen werden kann. Die Aktivierung von System 2 ist zudem ein bewusster Prozess. (vgl Kahneman 2011b)

Now this is disastrous in some ways. Because you can make a very coherent story, out of very little information, and out of information that is in fact not reliable. The quality of the story depends very little on the quality and quantity of the information. So people can be very confident - with very little reason. Confidence is therefore not very good diagnostic, for whether you can trust yourself, or somebody else. And If you are to evaluate weather you can trust somebody with a lot of confidence, that's not the way to do it, the way to do it, [...] is to ask - what environment have they been in, and have they had an opportunity to learn its regularities. Subjective confidence is not a good indicator. (Kahneman 2011b)

Kahnemann führt hier ein Argument im Sinne des post-positivistischen Wissenschaftsverständnisses.

Umgelegt auf das Beispiel mit dem Datenwissenschaftler der sich im Umgang mit dem Datensatz eine Expertise aufbaue und so dessen Verwertbarkeit durch die Prüfung entwickle, stellt sich hier die Frage in wie weit dabei, aufgrund des hohen Abstraktionsgrades, eine Einarbeitung in die „innere Ordnung“ (regularities) eines Umfeldes, im Sinne Kahnemans, möglich bleibt, die wiederum ein Expertenverständnis ermöglicht. Dies gilt sowohl für den Datenwissenschaftler der Datenbestände „reinholt und zusammenführt“, als auch für den Wissenschaftler der versucht aus einer signifikanten Korrelation heraus auf einen Kausalzusammenhang zu schließen. Gestaltet sich der jeweilige Versuch impulsgeleitet, lässt sich in jedem Fall auch aus der Sicht von Kahneman festhalten, dass das Vertrauen das dieser in eine derartige Expertise setzten würde, nicht besonders groß ausfällt. Dadurch ist dieser Aspekt auch hinsichtlich einer von Datenwissenschaftlern immer wieder referenzierten aktuellen wissenschaftlichen Betrachtungsgröße, welche auch den Prozess der Entscheidungsfindung selbst nicht ausspart, in Frage gestellt. Jedoch nicht ausschließlich im Bezug auf das „Versagen von Kausalitätsbehauptungen“, sondern auch bezüglich des Selektionsprozesses der Datenanalyse, wie auch im Bezug auf die Anwendung induktiver Logik bei einer versuchten Etablierung eines Kausalschlusses aus dem Korrelationsverstehen einer Big Data Analyse heraus. Cukier und Mayer-Schöneberger scheinen auch im Hinblick darauf gut beraten einen Kausalitätszwang für die Big Data Analyse abzulehnen.

Behavioral Economics stellt als Wissenschaftsrichtung explizit Kausalzusammenhänge in Frage und widerlegt solche regelmäßig in empirischen Forschungssettings – kann jedoch im Verständnis seiner Proponenten nicht in einer das Kausalitätsprinzip generell in seiner Existenzberechtigung hinterfragenden Ausrichtung verstanden werden. Die Tendenz von Proponenten des Big Data Ansatzes Experimentaldesigns der Behavioral Economics für Fallbeispiele heranzuziehen, relativiert sich vor diesem Hintergrund.

Ein weiterer häufig genannter Vertreter des Feldes, Dan Ariely – ehemaliger Professor am MIT, mit aktueller Professur für Psychologie und Behavioral Economics an der Duke University, umschreibt im Gegenzug dazu sein Verständnis von Big Data Analyse sogar pointiert wie folgt:

“Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it.” – (Ariely, 2013)

Woraus sich deutlich eine Kritik am Fehlen einer kausal begründeten inneren Logik ablesen lässt.

9.2 Gigerenzer

Gigerenzers Kritik an einem Prozess der sich anstatt eines klassischen sozialwissenschaftlichen Ansatzes, in dem zuerst der Inhalt eines Problems definiert wird, und erst darauffolgend die empirische Forschung beginnt, einem „praktischen Zusammenhang“ bedient, geht sogar noch über die Kahnemans, der zumindest einen Kausalschluss erwartet, hinaus.

Most practicing statisticians start by investigating the content of a problem, work out a set of assumptions, and, finally, build a statistical model based on these assumptions.

The heuristics-and-biases program starts at the opposite end. A convenient statistical principle, such as the conjunction rule or Bayes's rule, is chosen as normative, and some real-world content is filled in afterward, on the assumption that only structure matters. The content of the problem is not analyzed in building a normative model, nor are the specific assumptions people make about the situation. (Gigerenzer, 1996, S. 592)

Gigerenzer bezieht sich dabei nicht auf ein Beispiel aus der Big Data Welt, sondern kritisiert, in einer fachinternen Debatte, Kahnemans Definition von handlungsleitenden Heuristiken als zu wage, da diese post hoc auf viele Forschungsergebnisse projizierbar seien und gleichzeitig als zu spezifisch, da sie bestimmte verhaltensleitende Grundannahmen wie die konkrete Interpretation von Fragestellungen durch einzelne Individuen ausschließen würden.

Was bei der Anwendung auf das Big Data Paradigma nicht mehr in dieser Form kritisiert werden kann ist, dass der statistische Zusammenhang und die Beispiele für Entsprechungen in der „realen Welt“ zu Beginn des Prozesses separiert stehen würden.

Nicht desto weniger, lassen sich die beiden restlichen Kritikpunkte auch auf das Wissenschaftsverständnis von Big Data umlegen, gerade da es sich primär um eine Methodenkritik handelt. Der Inhalt des Problems wird (im Falle der Big Data Analyse) bei der Konstruktion des normativen Modells wenig beachtet, genausowenig wie spezifische Annahmen, die die Handlungssubjekte in einzelnen Situationen beeinflussen. Beide Aspekte finden sich auch in der Reaktion unterschiedlicher Forschungsrichtungen auf das Proposal von Bentley et al (2014).

Gigerenzer kritisiert zudem in direkter Folge ein erkenntnistheoretisches Problem unter Statistikern, wonach es nach den beiden populärsten, divergierenden, internen Richtungen nicht eindeutig sei, ob sich statistische Vorhersagen überhaupt auf Einzelereignisse beziehen können. Gigerenzer referenziert dazu auf einen seiner Artikel aus dem Jahr 1994 unter dem Titel „Why the distinction between single-event probabilities and frequencies is relevant for psychology and vice versa“.

In diesem beschreibt er das Problem wie folgt.

Does probability theory apply to single events? This is a controversial matter amongst probabilists, who have long and heatedly debated the merits of subjective Bayesian versus objective frequentist interpretations of probability. The influential Bayesian Leonard J. Savage (1954), for instance, introduced his notion of personal probability with every-day reasoning about singular events: "I personally consider it more probable that a Republican president will be elected in 1996 than that it will snow in Chicago sometime in the month of May, 1994. But even this late spring snow seems to me more probable than that Adolf Hitler is still alive." (p.27).

[...]

The mathematician Richard von Mises (1957) was one of those many [who disagreed]. [...] The number of deaths at age 41 was 940 out of 85,020, which corresponds to a relative frequency of about .011. This probability is attached to a class, but not to a particular person or event. Every particular person is always a member of many different classes, whose relative frequencies of death may have different values. Therefore, von Mises concluded: "It is utter nonsense to say, for instance, that Mr. X, now aged forty, has the probability 0.011 of dying in the course of the next year." (p.17-18). By now it should be clear that according to a strong frequency view of probability (e.g., Neyman, 1977; von Mises, 1957), what has been labeled the conjunction fallacy is not an error in probabilistic reasoning. In this view, probability theory is about frequencies and simply doesn't apply to single events. (Gigerenzer, 1994, S. 132)

Gigerenzer folgt weiters aus, dass die zweite Denkschule unter Statistikern an den statistischen Abteilungen von Universitäten weiter verbreitet sei.

Die Implikationen bezüglich des Big Data Paradigmas sind schwer verkennbar. Wird keine Konkretisierung hinsichtlich einer Korrelationsbeschreibung vorgenommen, die diese als „wiederholt auftretendes Ereignis“ fassen könne, wäre gege-

benenfalls keine Wahrscheinlichkeitsaussage mehr möglich. Zudem wäre das Extrapolieren von kumulierten Wahrscheinlichkeiten – um daraus Aussagen über ein Individuum zu treffen unmöglich.

Diese Sichtweise führt wieder deutlich zu der Auffassung, dass auch Big Data Ansätze nicht ohne eine Kausalbehauptung auskommen. Sie macht erstmals auch Mayer-Schönbergers Bestehen auf einer gesellschaftlichen Grundbedingung eines „Big Data Anwalts“ für das einzelne Datensubjekt, aus einem theoriegeleiteten Verständnis verständlich. Es ist aus der Theorie heraus argumentierbar, dass es nicht ausreicht gruppenspezifische Wahrscheinlichkeiten auf das Individuum aufzurechnen, wenn nicht gleichzeitig alle weiteren das Individuum in dieser Sache betreffenden Gruppenwahrscheinlichkeiten erfasst werden können. Handlungsansätze können vor dem Hinblick eines Gruppenphänomens gefordert werden, jedoch kann diesbezüglich nicht angenommen werden, dass sie auf jedes Datensubjekt einer Gruppe gerechtfertigt Anwendung finden.

10. Ethik

Unabhängig von der aktuellen theoretischen Verortbarkeit der Big Data Analyse, findet diese genauso wie die Aggregation von Datenbeständen über Gesellschaftsverhalten als Teil unseres täglichen Alltags Anwendung. Als solche ist sie auch Teil des aktuellen politischen Diskurses und prägt die Entscheidungsprozesse unserer Gesellschaftsentwicklung.

Beziehen wir uns rückblickend auf den Ausspruch Klimburgs, wonach Big Data der Rohstoff der aktuellen Informationsgesellschaft sei, und Externalitäten des Prozesses vermehrt ignoriert werden würden, wird unabhängig von einer wissenschaftstheoretischen Verortung des Phänomens, auch die Frage seiner gesellschaftlichen Bedeutung tragend. Insbesondere da mit Big Data ein Paradigmenwechseln in vielen bisher handlungsleitenden Vorstellungen einhergeht, soll dieser auch im Hinblick auf seinen ethischen Anspruch reflektiert werden.

10.1 Bewusste Partizipation und Wahrung der Privatsphäre

Eine der ersten Fragen die die Betrachtung der Ethik von Big Data Analyse aufwirft, ist die zur Partizipation. Ist es einem Individuum möglich sich einer Partizipation zu verweigern, wie stimmt es dieser zu, ist eine individuelle Abschätzung der Auswirkungen möglich, besteht ein möglicherweise ungerechtfertigter Zugriff auf die Privatsphäre. Das Beispiel das in diesem Zusammenhang immer wieder fällt ist eines das Charles Duhigg 2012 in einer Reportage der New York Times erstmals aufwirft.²³

Ein Mann betritt in Minneapolis eine Filiale des Discounters Target und verlangt den Manager zu sprechen. Dem legt er einen Zettel mit Rabattgutscheinen vor und verlangt lauthals eine Erklärung dafür, dass man seiner Tochter Werbung für „Baby-

²³ How Companies Learn Your Secrets, http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=0 [Letzter Aufruf: 19. 04. 2015];

kleidung und Krippen“ zukommen lasse, diese sei schließlich noch in der Highschool
- Direktzitat: „Versuchen sie sie zu ermuntern schwanger zu werden?“

Der Manager wirft einen Blick in seine Datenbank, sieht dass die betreffende Person tatsächlich für Zusendungen aus dem Bereich Mutterschaft vorgemerkt ist, streicht den Vermerk, und entschuldigt sich.

Als er wenige Tage später bei der Familie anruft, um sich erneut zu entschuldigen, wird ihm vom Vater der Tochter unterbreitet, dass er es sei, der sich zu entschuldigen habe, er hätte ein Gespräch mit seiner Tochter und anderen Familienmitgliedern geführt, seine Tochter wäre tatsächlich schwanger, er sei nicht über alle Vorgänge informiert gewesen.

Duhigg berichtet im selben Artikel darüber, dass das Unternehmen Target den Wirtschaftswissenschaftler und Statistiker Andrew Pole engagiert habe um aus dem Kaufverhalten seiner Kunden zu identifizieren ob diese schwanger sind. Genauer gesprochen im zweiten Trimester. Dieser Zeitraum wurde von dem Unternehmen als eines der wenigen Fenster identifiziert in dem sich das einmal geprägte Kaufverhalten eines Konsumenten (welche Geschäfte er besucht, wo er welche Waren kauft) noch einmal besonders effektiv ändern ließe. Pole bestätigt gegenüber Duhigg, dass er und sein Team diesbezüglich erfolgreich waren. Bestimmte Produktkategorien (parfümfreie Körperlotionen, Nahrungsergänzungspräparate, ...) konnten als Marker für die bevorstehende Geburt eines Kindes identifiziert werden.

Darauf angesprochen, nahm das Unternehmen selbst gegenüber Duhigg nicht Stellung.

Ein zweites Beispiel das an dieser Stelle noch einmal Mal Erwähnung finden muss, ist die Datensammlungspraxis von Geheimdiensten im Rahmen dessen was diese in Deutschland intern als „Weltraumtheorie“ bezeichnen. Das Prinzip, dass wer an bestimmten Knotenpunkten im Ausland, oder sogar direkt von Satelliten im Weltall, Massendaten abschöpft, gegen keinerlei Gesetze innerhalb des eigenen Landes verstößt. Diese gängige Praxis innerhalb eines rechtlichen Graubereichs soll nun auch

durch ein sich bereits im Entwurf befindliches Gesetzesvorhaben legitimiert werden.²⁴

Alle im ersten Absatz genannten Problemfelder zur bewussten Partizipation werden von diesen beiden Beispielen berührt, wobei keiner der genannten Aspekte im Interesse der Wahrung der Privatsphäre von Individuen umgesetzt wurde. Der erste betrachtete Fall verdeutlicht auch die Schwierigkeit auf technische Lösungen zum „Opt Out“ („aus dem Verteiler nehmen“) zu setzen, da die individuelle Ausprägung und Bedeutung der Auswirkungen teilweise nicht einmal antizipierbar wäre.

10.2 Datenschutz, ein Ausblick

Dass vor dem Hintergrund der neuen Entwicklungen auf dem Feld der Datensammlung und Auswertung ein neues Reglement im Hinblick auf den Datenschutz, auf den Weg gebracht werden soll, ist ein aufgeklärtes Ansinnen. Die Europäische Union hat vor diesem Hintergrund den Prozess einer Reform ihrer Datenschutzdirektive eingeleitet und 2012 die ersten Änderungsvorschläge zu Papier gebracht.

Wenn wir in der Folge die Vorlagen und Änderungen betrachten, die sich im Laufe des Gesetzwerdungsprozesses noch weiter ändern werden, dann immer unter dem Vorbehalt, dass die entsprechenden Entwürfe in der Anfangsphase als „Leaks“, als nichtgenehmigte und anonyme Veröffentlichungen, an das Licht der Öffentlichkeit gelangten.

Mittlerweile hat das Paket den Ministerrat passiert, und bezüglich der proklamierten moralischen Imperative und Auflagen, die auch in der Außenwahrnehmung durchaus auch maßgeblich das Bild Europas prägen würden - im historischen Zusammenhang, wie auch in der Hoffnung, dass zumindest Europa sich verstärkt seinen Bürgerrechten verpflichtet sieht – wurden starke Konzessionen hingenommen.

²⁴ vgl. <http://www.spiegel.de/netzwelt/netzpolitik/bundesregierung-plant-gesetz-zur-spionage-im-weltraum-a-1027963.html> [Letzter Aufruf: 19. 04. 2015];

Ich zitiere teilsinhaltlich den Journalisten Moechel im Mai 2013.

„EU-Ministerrat demontiert Datenschutzpaket - Österreich hat wegen der starken Veränderungen des Parlamentsentwurfs im Ministerrat generellen Vorbehalt gegen die ersten 37 Artikel insgesamt eingelegt.“ (Moechel, 2013)

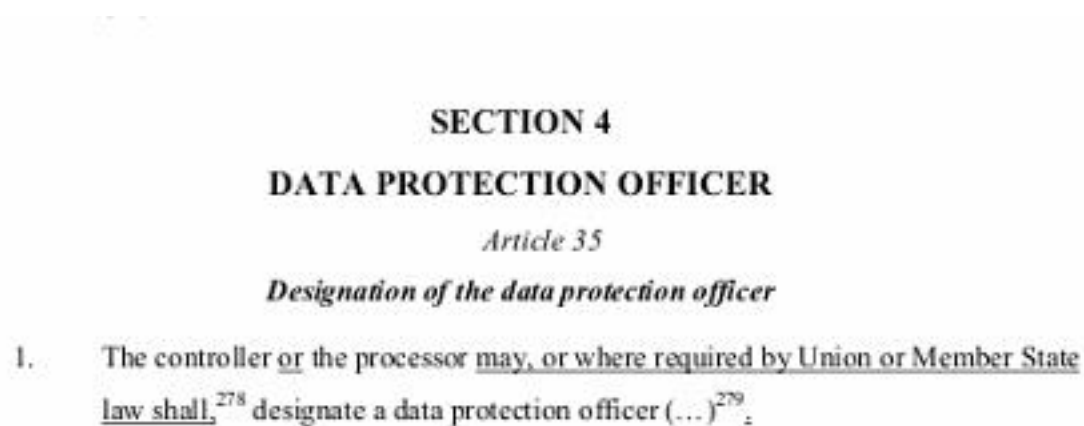


Abb. 5: Einleitung der Sektion 4 des neuen Entwurfs der Datenschutzdirektive. Inhaltlich veränderte Stelle durch Unterstreichen hervorgehoben.

Moechel verweist dazu exemplarisch auf Sektion 4 des Entwurfs in dem es unter Artikel 35 zu den Aufgaben von Datenschutzbeauftragten in Daten aggregierenden Unternehmen nunmehr heißt:

„Der Inhaber oder der Verwerter kann, oder wo von der Union oder einzelnen Mitgliedstaaten vorgesehen, soll einen Datenschutzbeauftragten einsetzen.“

Die Kritik des Artikels im folgenden, wie die generelle Kritik an den Änderungen der Vorlage, bezieht sich in Summe auf einen erkennbaren Trend der aus gesetzlichen Vorgaben - „kann und soll Bestimmungen“ werden hat lassen, aus zum Teil konkret bezifferten Strafzahlungen - „Verwarnungen“ und aus rechtlichen Vorgaben - „Selbstverpflichtungen mittels Codes of Conduct“ (von Unternehmen selbst konstituierte „Verhaltensregeln“).

Ein derart umfassendes Fehlen an Standardisierungsvorgaben (kann - , soll - Bestimmungen) ist ungeeignet dem Anspruch bereits heute vollzogener Praxis, im Bürgerinteresse entgegenzuwirken, zu entsprechen.

Dabei richtet sich die Kritik nicht sosehr darauf, auf Selbstverpflichtungen von Unternehmen abzustellen, sondern die Auswirkungen bei Nichteinhaltung von Prinzipien der bisherigen Bestimmung sukzessive zurückzufahren.

In diesem Zusammenhang sei auch auf die Zahl von 3.133 Änderungsanträgen zum Stand März 2013 verwiesen²⁵, die hier auf EU-Ebene im Zuge des Gesetzwerdungsprozesses binnen eines Jahres eingegangen sind. Im direkten Vergleich, selbst die ebenfalls im Konzept konsumentenorientierte EU-Richtlinie zur Lebensmittelkennzeichnung brachte es im Zeitraum von zwei Jahren vor ihrer Einführung auf vergleichsweise geringe 1500 Änderungsanträge.²⁶

10.3 Verteilung von monetärem Wert

Eine weitere Frage die unter ethischen Gesichtspunkten betrachtet werden muss ist, wie der aus Big Data generierte Wert verteilt ist.

Big Data generiert Wert. So existiert ein Bericht des World Economic Forums aus 2011 unter dem Titel „Personal Data: The Emergence of a New Asset Class“ (vgl. Schwab et al., 2011), bei dem Überlegungen angestellt werden, wie Big Data – hier unter der Bezeichnung Personal Data, jedoch vor dem Hintergrund der Begriffsdefinition von Big Data in dieser Arbeit vollkommen inhaltsgleich – zukünftig als ökonomische Wertkategorie begriffen und in ökonomischen Modellen Verwendung finden könnte.

Gleichzeitig wirft der Österreicher Christian Fuchs, Professor für Social Media Studies an der Universität Westminster, ein Problem auf, das er als „New media and class exploitation“ umschreibt. (vgl. Fuchs 2013)

²⁵ vgl. <http://derstandard.at/1362108027424/3133-Aenderungsantraege-fuer-EU-Datenschutz-Reform>

²⁶ vgl. <http://www.swp.de/ulm/nachrichten/politik/Mehr-als-1500-Aenderungsantraege;art4306,406656>

Nutzer von sozialen Medienplattformen die beginnen nutzergenerierte Inhalte zu verfassen und diese mit anderen zu teilen, produzieren nach klassischen ökonomischen Theorien Mehrwert den sie dem jeweiligen Unternehmen überantworten, werden dafür jedoch nicht monetär entlohnt. Den Nutzern werde dabei kein Produkt verkauft, auch nicht ein Service im klassischen Sinn, das Unternehmen jedoch verkauft den jeweiligen Inhalt und die daran gebundene Aufmerksamkeit als Ware an Werbeunternehmen weiter.

Die Nutzerakquise verlaufe dabei nach dem Modell bestehender Massenmedien- jedoch bestehe im Gegensatz zu diesen der Unterschied, dass Nutzer selbst aktiv die Inhalte generieren. Auch andere Aktivitäten wie Kommunikation, oder Community-Building würden von den Unternehmen als Mehrwert genutzt, verarbeitet und verkauft.

Die Gewinner dieses Prozesses seien nach Ekbia et al. (2014) eine Gruppe an „nouveau riche“ die solche Plattformen aufbauen und nachdem sich um sie erfolgreich eine aktive Community (Nutzerschaft) gebildet hat mehrheitlich wieder verkaufen, wobei der Großteil des Wertes eines solchen Unternehmens von unbezahlten Nutzern generiert worden sei.

Ekbia et al. bezeichnen die Nutzer als Verlierer dieses Prozesses, die die Rolle eines Stand-Ins für die erste Gruppe spielen müssen, die „Links“ innerhalb dieses Netzwerks erstellen würden, die das Netzwerk aktiv, produktiv und nützlich halten und von der ersten Gruppe in einer auf bestimmten Erwartungshaltungen beruhenden Organisationsform in das Netzwerk einbinden, in der große Löcher und Lücken klaffen würden, von denen erwartet werden würde, dass sie die Nutzer füllen.

Vor allem vor dem Hintergrund, dass sich einige der größten aktuellen Datensammlungen bei Unternehmen befinden die dieser Ausrichtung gefolgt sind, verstärkt den Stellenwert dieser Betrachtung.

Auch dass einige Unternehmen, darunter Google Inc., damit begonnen haben Gewinnanteile an die Ersteller von Inhalten weiterzureichen, verdeutlicht den Zwiespalt. Der „Youtuber“ ist heute zu einer mitunter belächelten Berufsbezeichnung

geworden. Ohne Marktkonkurrenz für seinen Arbeitgeber die ihn abwerben würde und ohne Gewerkschaft im Rücken.

Sprechen wir von rein automatisiert generierten Daten, wobei sich beispielsweise die oben genannten sozialen Verknüpfungen von Datenpunkten explizit als Teil einer Bedeutungszuweisung verstehen lassen, die als Metadaten in ein Dataset eingehen, so haben auch diese einen Wert der selten abgegolten wird.

10.4 Das soziale Selbstbild

Eine sehr interessante Entwicklung die ebenfalls von Ebkia et al. (2014) angesprochen wird, besteht im vorherrschenden Selbstbild des datengetriebenen „Neuen Menschen“. Jede Entwicklungsstufe unserer Gesellschaft habe auch das ideale Selbstbild seiner Bürger geprägt – das aktuell vorherrschende schreibe ihm folgende idealisierte Werte zu: „Hohe Unabhängigkeit, stark involviert, selbstständig, sehr mobil, mit dem Potential um-lernen zu können, seine Fähigkeiten neu zu entwerfen, und ortsungebunden zu sein“. Der neue ideale Mensch ziehe seinen Stolz daraus „Abhängigkeiten zu meiden und gleichzeitig sein eigener medizinischer Berater und Pensionsfondmanager“ zu sein. Dieses Bild wäre durch die neue Datenwelt geprägt und würde insbesondere über soziale Medien Verbreitung finden. Ekbia et al. attestieren sogar social engineering indem Big Data das Menschenbild über drei Aspekte forme. Einen „Drei-Ebenen-Mechanismus“ sozialer Kontrolle.

(Verhaltens-)monitoring, (Verhaltens-)mining und (Verhaltens-)manipulation

Dies ließe die Durchschnittsperson mit einem ambivalenten Gefühl von Ermächtigung und emanzipatorischem Selbstausdrucksvermögen zurück, kombiniert mit Besorgnis und Verwirrtheit. (vgl. Ebkia et al. 2014, S. 5f)

Auf der Suche nach einem beispielhaften Ausdruck eines solchen erlebten Zwierspalts werden wir erneut bei Haddow (Adbusters Magazine) fündig. (vgl. Haddow 2014)

Der Autor beschreibt darin ein Vorstellungsgespräch das in einem von der jungen Startup Szene frequentierten Pub, innerhalb eines trendigen, technologieaffinen Stadtteils in Vancouver stattfindet – bei dem er sich bei einem Unternehmen bewirbt, das sich auf Online Brand-Management Systeme spezialisiert hat (*„Which is a euphemism for a human centipede of marketers selling marketing to marketers for marketing. The firm is worth a billion dollars.“*).

Das Bewerbungsgespräch läuft nicht sehr gut, und während er sich in dem Lokal umsieht oder in der Abwesenheit des Human Resources Managers am Handy auf seiner Facebook Seite vorbeisurft (*„I load up Facebook in the interim, hoping to find a shard of inspiration in my feed that will provide a topical talking point. Instead I find a listicle. A curiosity gap headline. An ad. A solicitation. Another ad. Another listicle. Oh dear, someone has lost their phone. And finally, an ad in the form of a listicle. Or is it a listicle in the form of an ad?“*) reflektiert er bezüglich dem was das Internet für ihn einmal war – und was daraus aus seiner Sicht geworden ist.

Ausgehend von einem Zitat von Fred Turner, einem Kommunikationswissenschaftler an der Universität Stanford, der die Ursprünge der heutigen Technologieszene in der amerikanischen Gegenkultur der 1960er sieht - diese Kultur die das Internet als einen Gegenentwurf zur bisherigen Gesellschaft sah, habe das idealisierte Verständnis von dem was eine Virtuelle Gesellschaft (community) und ein digitaler Bürger sein kann nachhaltig geprägt – entwirft Haddow eine historische Entwicklung, exemplarisch am Beispiel von Steward Brand.

Steward Brand war 1968 Herausgeber des Whole Earth Catalog, einem Kompendium von Literatur der Gegenkultur bis 1971, finanzierte am Tag der Auflösung des Magazins die Projekte von Fred Moore, der den Homebrew Computer Club gründete aus dem in indirekter Folge auch Apple Inc. hervorging. 1985 gründete er The Well (Whole Earth 'Lectronic Link) eine der ersten virtuellen Gemeinschaften in der Geschichte des Webs, 1987 das Global Business Network (GBN), einen Think Tank der als einer der ersten galt die langfristige Planungsmodelle popularisierten.

Das GBN beriet später Royal Dutch Shell und AT&T zum Theoriekomplex „Bewältigung sozialer Unsicherheit“ und wurde 2000 von der Unternehmensberatung Monitor Group aufgekauft. Monitor Group kam 2011 in die Schlagzeilen als bekannt wurde, dass das Unternehmen einen hochdotierten Auftrag angenommen hatte, Muammar Gaddafi's Öffentlichkeitsbild im Westen zu verbessern.²⁷

Dieses historische Beispiel zeigt den Zwiespalt zwischen den Idealen einer Gegenkultur (Freiheit, Unabhängigkeit, Selbstständigkeit, der Fähigkeit sich neu zu entwerfen) und einem vergleichsweise inhaltsleeren Pragmatismus wie er auch von Ebkia et al. skizziert wird.

Haddow fasst für sich zusammen:

„Das Internet ist eine gescheiterte Utopie. Und wir sind alle in ihr gefangen. Aber ich bin noch nicht bereit aufzugeben. Es war wo ich das erste Mal Punk Rock und Anarchismus entdeckt habe. Wo ich von Yi Jing und Albert Camus erfahren habe, während ich „Holiday in Cambodia“ [Anm. Song der Dead Kennedys, kalifornische Punk Band]“ mit 15 kbps heruntergeladen habe. Es ist wo ich das erste Mal über den Photos eines Mädchens pervers wurde (perved out) in die ich mich später verlieben sollte. Es ist Heimat für mich, dich und alle die wir kennen.“ (Haddow 2014)

Auch wenn es sich hierbei nur um eine emphatische Einzelbetrachtung handelt, gefärbt und fatalistisch, die Ethik hinter dem Menschenbild das aus dem Internet, dem quantifizierbaren Menschen und einer vernetzten Gesellschaft heraus entstanden ist, ist eine eigene Betrachtung wert. Die nähere Beschäftigung damit geht jedoch über das hinaus, was diese Arbeit leisten kann.

²⁷ siehe ebd., sowie „US firm Monitor Group admits mistakes over \$3m Gaddafi deal“, <http://www.theguardian.com/world/2011/mar/04/monitor-group-us-libya-gaddafi> Datum des letzten Aufrufs: 11. 04. 2015;

10.5 Recht auf Vergessenwerden

Das bereits eingangs erwähnte „Recht auf Vergessenwerden“ von Mayer-Schönberger beschreibt ebenfalls eine neue ethische Grundabwägung die zu einer gesellschaftspolitischen Notwendigkeit wird.

Mayer-Schönberger illustriert dies am Beispiel des Psychotherapeuten Andrew Feldmar der in den 1960er Jahren LSD konsumiert hatte und 2001 in einem von ihm verfassten Artikel in einem interdisziplinären Journal darüber schreibt. Der Artikel geht online und wird 2006 von einem amerikanischen Grenzbeamten gefunden der Feldmar der, wie bereits mehrere Male zuvor, aus Canada anreist um einen Freund vom Flughafen abzuholen, anhand seines Namens googlet. Der Grenzbeamte besteht darauf, dass Feldmar den Umstand seines ehemaligen Drogenkonsums von sich aus und ungefragt bei der Einreise angeben hätte müssen und verweigert ihm die Weiterfahrt. In weiterer Folge wird ein generelles Einreiseverbot über ihn verhängt, das bis heute aufrecht ist. (vgl. Mayer-Schönberger 2010, S. 4f)

Mayer-Schönberger spricht in dem Zusammenhang davon, dass er der letzten Generation angehören würde die noch „vergessen werden kann“ und dass sich derartig verändert habende Grundbedingungen von dem was einen Menschen ausmacht, tiefgreifende gesellschaftliche Veränderungen in sich bergen. Eine Annahme von „Unbeobachtetheit“ sei im heutigen Big Data Paradigma nicht mehr gegeben, auch sei der Zeitbegriff im Zusammenhang mit dem Leben eines Einzelnen oftmals nicht mehr zyklisch definiert, sondern eine kontinuierliche Linie geworden, in der jeder Zeit nach vor oder zurück geblättert werden kann und das mit absoluter Präzision.

Dies ersetze wesentliche Teile einer Annahme von Privatsphäre vollständig. Wird ein „Recht auf Vergessenwerden“ postuliert, eröffnet es wie eingangs bereits an einer Auseinandersetzung zwischen Google Inc. und dem EuGH verdeutlicht ein Konfliktfeld zwischen legitimen Interessen des Einzelnen und dem Recht auf Informationsfreiheit einer Gesellschaft. Diese Abwägung dem Unternehmen das den Suchindex stellt selbst zu überlassen, bleibt bestenfalls ein Kompromiss.

Google Inc. ist bemüht in dieser Frage auch öffentlich zu agieren und hat einen Kriterienkatalog in der Zusammenarbeit mit dem von ihnen eingesetzten Expertengremium erstellt, der nun auch offen zugänglich ist.

Darin wird ein Mehrkriteriensystem beschrieben in dem die Fragen ob die betreffende Person eine Person des öffentlichen Lebens ist, ob sie einer Veröffentlichung der Information zugestimmt habe und wie lange die Veröffentlichung zurückliegt - Einfluss auf die Erfolgswahrscheinlichkeiten eines Löschungsantrages haben.

Google Inc. gibt jedoch weiters auch inhaltsspezifische Kriterien an, bei denen eine Löschung als unwahrscheinlich gelten müsse. Diese betreffen beispielsweise Texte mit politischem, religiösem oder philosophischem Inhalt. Korrekte Informationen ohne Drohpotential, historisch relevante Informationen und wissenschaftliche, sowie künstlerische Beiträge. (vgl. Demgen 2015)

Der von Mayer-Schönberger oben zitierte Fall hätte also vermutlich kaum eine Chance auf Tilgung und wäre von der Umsetzung der Prinzips eines „Rechts auf Vergessenwerdens“ nicht berührt.

Google Inc. nennt auch „Erfahrungsberichte von Verbrauchern“ als ein inhaltliches Kriterium das eine Löschung unwahrscheinlich mache. Der ethische Sinn dieser Eingrenzung eröffnet sich nicht zwangsläufig, tatsächlich stärkt er die Sichtweise, dass auch Unternehmensinteressen die Wahrscheinlichkeit einer Entfernung aus dem Suchindex berühren. Weiters ist der tatsächliche Entscheidungsprozess in letzter Konsequenz nicht offen einsehbar und vor dem Hinblick der Anzahl an zu bearbeitenden Vorgängen in weiteren Teilen automatisiert. So erreichten Google Inc. laut einer Angabe in einem Die Presse Artikel binnen der ersten vier Monate nach Start des Programms 146.000 Löschanträge von denen nach Unternehmensangaben 41,8% stattgegeben wurde. (vgl. Grech, 2014)

Das Recht auf Vergessenwerden ist auch innerhalb anderer Unternehmen die hier an der Speerspitze der Entwicklung stehen ein Thema. So existiert die Technologie der Gesichtserkennung auf Bildern bereits seit mehreren Jahren, mit Facebook Inc.

als Besitzer einer der diesbezüglich größten kommerziellen, biometrischen Datenbanken. „Sag uns wer noch auf diesem Photo zu sehen ist“ war der Leitspruch einer Generation an Facebook Nutzern die andere aus sozialen Incentives heraus auf Photos identifizierte. Seit Februar 2013 wert Facebook entsprechende Daten auf Nutzerbildern im Hinblick auf eine europäische Richtlinie nicht mehr aus. Die entsprechenden bereits gesammelten Daten seien gelöscht worden.²⁸

10.6 Anonymisiert und nicht anonym

Ein Problem stellt sich hier bereits vor dem Hintergrund der Aggregation von Daten. Wie von Klimburg im ersten Teil dieser Arbeit skizziert, basiert ein Großteil der aktuellen Ökonomie von Social Media Portalen und Web Angeboten auf der Aggregation von Daten die zum Teil von den Plattformen selbst benutzt, zum anderen Teil, meist anonymisiert, weiterverkauft werden. Dabei ist es von hohem ethischen Stellenwert, ob sich aus einem Datum das nicht mehr durch die Nennung eines Namens oder eines anderen einzigartigen Identifizierungsmerkmals als persönlich zuordenbar gilt, durch die rückbeziehende Verknüpfung anderer, einsehbarer Merkmale mit zusätzlichem Material, wieder Rückschlüsse auf die Person zulassen die eine erneute Identifizierung (Deanonymisierung) ermöglichen.

Auch im wissenschaftlichen Umfeld ist man nicht vor dieser Problematik gefeit. 2009 veröffentlichte eine Gruppe von Forschern des Berkman Centers für Internet und Gesellschaft in Harvard ein anonymisiertes Dataset von mehreren tausend facebook Profilen einer „anonymen Universität im Nordosten Amerikas“ die von mehreren Forschungsassistenten per einloggen auf Facebook und kopieren der Informationen auf verfügbaren Profilen aggregiert wurde. Abhängig vom jeweiligen Forschungsassistenten ergaben sich daraus auch unterschiedliche Profilzugriffsrechte und daraus folgend eine höhere Detailtiefe in einzelnen Profilinformatoren. Weiters wurden die jeweiligen Studienschwerpunkte der einzelnen Studenten nicht redigiert, was in einer Zusammenführung der etwas weniger geläufigen eine eindeuti-

²⁸ vgl. <http://www.zeit.de/digital/datenschutz/2013-02/facebook-gesichtserkennung-verfahren-caspar> [letzter Aufruf: 20. 04. 2015]

ge Zuordnung der Universität erlaubte. Wenig überraschend die selbige aus der auch die Wissenschaftler heraus publizierten.

Erschwerend kam hinzu, dass von keinem der Proponenten der Studie zuvor ein Einverständnis eingeholt wurde. (vgl. Zimmer 2008)

Das Dataset wurde nach öffentlich werden der Problematik depubliziert, und findet sich trotz einem entsprechenden Eintrag, dass man sich um die weitere Anonymisierung der Proponenten kümmere bis heute nicht mehr im öffentliche einsehbaren Datenbestand der Universität.²⁹

Aus dieser Veröffentlichung heraus entstand eine neue Diskussion über die Validität von Anonymisierungsgrundsätzen innerhalb der Wissenschaften, wobei von Ohm (2010) vor allem das „Veröffentlichen und Vergessen“ angeprangert wird. Bei diesem werden vordergründig eindeutige Felder mit persönlich identifizierbarer Information entfernt, oder im Sinne einer noch Auswertbarkeit im Rahmen der Studie generalisiert. Ohm zeigt anhand von drei Beispielen, wie durch weniger eindeutige Identifier in Kombination mit der statistischen Verknüpfbarkeit der Daten mit anderen Informationen wieder eine sehr wahrscheinliche bis eindeutige Reidentifizierung möglich machen.

Ohm referenziert darunter eine Studie aus dem Jahr 2010³⁰ aus der hervorgeht, das ausgehend von den Volksbefragungsdaten des US Census aus dem Jahr 1990, 87.1 Prozent der Gesamtbevölkerung der USA durch die Kombination von Postleitzahl, Geburtstag und Geschlecht eindeutig identifizierbar waren.

Dabei sei es nicht notwendig, dass sich alle zu einer Reidentifizierung notwendigen Datenpunkte im selben Dataset befinden, wenn aufgrund bestimmter Muster eine Zuordnung zu anderen Datenbeständen möglich wird. In einem zweiten von Ohm skizzierten Beispiel, einer Veröffentlichung eines Datasets mit hundert Millionen Einträgen des Online Streaminganbieters Netflix, war es möglich scheinbar anonymisierte Nutzer ausgehend von ihrer Bewertungs Historie, offen zugänglichen Profi-

²⁹ Siehe <http://thedata.harvard.edu/dvn/dv/t3> [letzter Aufruf: 14.04.2015]

³⁰ vgl. Sweeney (2010)

len auf der Webseite zuzuordnen aus denen wieder andere Identifier, wie eine eMail Adresse oder ein Name erfahrbar wurden. Es reichte ein Sample aus sechs weniger populären Filmen um eine Person mit einer Wahrscheinlichkeit von 84% eindeutig zu identifizieren. Nahm man ein Sample aus beliebigen sechs Filmen und kannte man die zeitliche Periode in der die Bewertung erfolgte (in einem Genauigkeitsintervall von 2 Wochen) war eine eindeutige Zuordenbarkeit mit 99% Wahrscheinlichkeit gegeben.

Ohm kommt zu dem Schluss, dass wir uns in einem Post-Anonymisierungs Zeitalter befinden in dem der strukturelle Umgang mit Datasets (Risikofaktoren bestimmter Identifier, Veröffentlichung oder nicht, Menge der veröffentlichten Daten, das Motiv vor dem sie veröffentlicht werden müssen, ...) wichtiger wird als alleine ihre Anonymisierung.

Ein Teil davon lässt sich unter den Grundsätzen von Datenvermeidung und Datensparsamkeit subsummieren, zwei zentralen Aspekten des deutschen Datenschutzrechts.

Im März diesen Jahres berichtet die Onlineplattform heise.de von einem weiteren Vorstoß der EU Kommission (Interinstitutional File: 0011 (COD)) diesen Prinzipien zumindest in ihrer absoluten Form nicht zu folgen.

Firmen, öffentliche Verwaltungen und sogar "Drittparteien" sollen dem Beschluss nach persönliche Informationen schon dann für andere Zwecke als ursprünglich angegeben verarbeiten dürfen, wenn ihre "legitimen Interessen" schwerer wiegen als die der Betroffenen. [...] Vorgaben zum sparsamen Umgang mit personenbezogenen Informationen im Sinne des etwa im Bundesdatenschutzgesetz noch festgeschriebenen Ansatzes der "Datenvermeidung" haben die Regierungsvertreter aus dem Entwurf gestrichen. Sie folgten damit der Empfehlung der federführenden Ratsarbeitsgruppe. Woher der Wind bei dieser Abkehr von bisherigen Datenschutzprinzipien weht, zeigt ein neu eingefügter Erwägungsgrund. Darin unterstreichen die Minister im Einklang mit einem Vorstoß der früheren italienischen Ratspräsidentschaft: Das

Verarbeiten persönlicher Daten für Zwecke des Direktmarketings soll von den berechtigten Interessen der Informationssammler gedeckt sein. (Krempf, 2015)

11. Fazit

Eine methodologische Kompatibilität des Big Data Ansatzes mit dem aktuell vorherrschenden Wissenschaftsverständnis ist theoretisch gegeben. Genauso schwach und unenthusiastisch wie diese Aussage fällt, ist sie auch zu verstehen, denn mit weiterhin umstrittenen Detailbetrachtungen, etwa in dem Bereich ob es sich beim Setzen von behaupteten Kausalschlüssen (und seien sie auch nur impliziert) auf der Basis von Zusammenhängen die aus dem Datenmaterial deutlich werden, vermehrt um induktive Logik handelt und in wie weit die Anwendung einer solchen notwendig, oder noch gedeckt ist, steht und fällt dieser Logikschluss.

Wenn Big Data Proponenten darauf verweisen, dass die subjektive Wertung die durch deduktive Forschungsansätze ausgeschlossen werden soll, ab einem bestimmten Umfang der Datenmenge nicht mehr gegeben sei, so bleibt dem entgegenzusetzen, dass bei der Aufbereitung des Datenmaterials (Sampling und Cleaning) immer noch ein starker subjektiver Einfluss besteht und induktive Logik am Ende einer Big Data Analyse von Scheinzusammenhängen nicht gefeit ist. Gerade Popper ist es jedoch, der einer „größeren Theorie“ durchaus etwas abgewinnen kann, selbst wenn diese Unschärfen aufweist, und auch Kuhn führt als eines seiner Qualitätskriterien für Theorien den Anspruch eines großen Gültigkeitsbereiches auf. Letztlich steht es auch immer noch jedem Wissenschaftler frei eine Big Data gestützte Theorie in ihrer Wirkungsfähigkeit zu falsifizieren und gegebenenfalls sogar sein Paradigma zu wechseln.

Sobald die Diskussion jedoch in Sphären vorstößt wonach der Beleg einer Kausalfolge in größeren Zusammenhängen von so vielen Faktoren bestimmt sei, dass sie bei komplexeren Fragen nie von einem Menschen verstanden oder gezogen werden könnten, ihre Nutzung jedoch opportun und notwendig sei - ergeben sich daraus

explizite gesellschaftliche Problemstellungen die die Idee von Individualität, Selbstbestimmung und Wissenschaft im Kern berühren. Und gegebenenfalls nicht einmal im Selbstverständnis der Statistik unumstritten sind.

Ein Kernbereich der Big Data Analyse, nämlich die Voraussage von Entwicklungen auf Basis vergangener Gerichtetheit wird sehr schnell zu einem gesellschaftsrelevanten Bereich, wenn das Verständnis davon eine dogmatische Note erhält – da sich diese im Einzelfall bis in den konkreten Anwendungsbereich fortsetzen kann. Änderungen daran werden in weiterer Folge zur Aufgabe von Gesellschaftsprozessen, was auch ein wenig das Machtgefälle beschreibt in dem die gesamte Diskussion zu verorten ist.

Dazu kommt, dass nach dem aktuellem Verständnis der Umfang der Datenmenge im Hinblick auf das Auffinden stabiler Zusammenhänge (ein Teil des internen Qualitätsverständnisses von Big Data) fast wichtiger ist, als die Qualität des Algorithmus als Recherchetool, was systemisch Zentralisierungstendenzen verstärkt.

Was die Annäherung einiger Fürsprecher beider Seiten an neuere wissenschaftstheoretische Konzepte wie aus dem Feld der Behavioral Economics betrifft, so dürfte diese nicht in einer besonders hohen theoretischen Kompatibilität der einen, oder anderen Seite begründet sein.

Ansätze die Sozialpsychologie und die Sozialwissenschaften generell an das Verständnis der Datenwissenschaften vor dem Hintergrund dieses Themas aneinander heranzuführen sind aktuell vorhanden und von beidseitigem Bemühen geprägt. Neue Entwicklungen werden nicht verleugnet, wenngleich auch die Schärfe des Konfliktpotentials anhand einzelner Reaktionen deutlich wird.

Der breite Wunsch sich die gesellschaftlichen Implikationen und konkreten Anwendungsformen durch ein Expertenverständnis auch vor einem gesellschaftlichen Hintergrund legitimieren zu lassen scheint aktuell noch nicht verwirklicht, ein breiter interdisziplinärer Konsens ist selbst auf dem Niveau führender Vertreter der Forschungsrichtungen nicht gegeben, Detailfragen bezüglich einer theoretischen Kom-

patibilität bleiben weiterhin ungeklärt. Eine Risikoabwägung ist schwer zu leisten und in Ethikkommissionen wie sie in anderen datengeleiteten Forschungsfeldern bei der Legitimation von Forschungsansätzen Verwendung finden, fehlt möglicherweise das theoretische und methodische Verständnis. Bereits in der heutigen Verwendung ist abzulesen, dass unerwünschte Effekte erst im Einsatz selbst sichtbar werden und den Einzelnen oft erst nach mehreren Jahren berühren können.

Ansätze darauf zu reagieren, setzen auf einer institutionellen oder Prozessebene an, da es für das Individuum selbst nicht mehr zumutbar zu sein scheint, die Auswirkungen seines/ihres Digitalen Fußabdruckes zu überblicken, geschweige denn im Einzelfall seine Nutzung zu legitimieren oder diese zu verwalten. Während der Ansatz selbst durchaus streitbar bleiben kann, ist es der Modus seiner Umsetzbarkeit weniger. Dass eine solche Verwaltungsebene nicht unabhängig von den bisherigen „Datenriesen“ eingezogen wird, sondern in Form von Selbstverpflichtungen, oder Gesetzen praxisrelevant zu werden scheint, scheint bereits daraus legitimiert, dass eine solche Verwaltungsebene selbst zu einem Big Data Akteur aufsteigen würde. Interessenskonflikte inkludiert.

Zumindest in soweit, und für jedes konkrete Individuum westlicher Gesellschaften relevant, hat sich der dogmatische Ansatz von Big Data als „in seinen Auswirkungen nicht mehr individuell begreifbar“, bereits durchgesetzt.

Die Frage ob der behavioristische Begriff der Black Box unter dem Paradigma der Big Data Analyse ein für alle mal ausgeschaltet sei, kann dezidiert verneint werden, denn selbst wenn sich, bezogen auf das Individuum eine neue Diskussion über eine Voraussagbarkeit von Entscheidungen ergeben hat, die durchaus bereits Aspekte wie das Bekenntnis zum freien Willen an sich berührt, sind diese Entscheidungsannahmen vornehmlich noch nicht kausal begründet – noch weniger in einer Weise in der sie gesellschaftstheoretische Vorgänge umschreiben. Das Fehlen dieser Begründung macht diese schwer in einem breiten wissenschaftlichen Diskurs verhandelbar. Die Kritik diesbezüglich („Datenfundamentalismus“) kommt nicht nur aus den Sozialwissenschaften, sondern auch von Vertretern der Big Data Analyse selbst. Es kann argumentiert werden, dass sich die behavioristische Black Box letztlich sogar

noch vergrößert hat, da immer breitere gesellschaftsrelevante Zusammenhänge erfasst, erkannt, aber selten kausal begründet werden.

Dennoch, wir befinden uns hier am Beginn einer Entwicklung.

Positiv bleibt anzumerken, dass der wissenschaftliche Diskurs zwischen den Forschungsrichtungen zurzeit offen stattfindet und Tendenzen attestierbar sind eine Vereinbarkeit der wissenschaftstheoretischen Linien herbeizuführen. Jedoch scheint die konkrete Anwendung von Big Data Applikationen einem diesbezüglichen Konsens zuvorzukommen.

Abschließend sei angemerkt, dass die sozialwissenschaftliche Verortung des Big Data Paradigmas von großem gesellschaftlichem Interesse bleibt und von allen Wissenschaftsrichtungen als werthaltig angesehen wird. Zusätzliche Forschung und Debatte auf dem Gebiet sind notwendig, da die Bedeutung von Big Data Applikationen, die bereits heute die Ökonomie der Informationsgesellschaft nachhaltig prägt, auch in unserem täglichen Leben, nicht geringer werden wird.

Quellenverzeichnis

- Agrawal, D. & Das, S. & El Abbadi, A. (2011). Big data and cloud computing: current state and future opportunities, Proceedings of the 14th International Conference on Extending Database Technology, New York: ACM;
- Analytis, P. P. & Moussaïd, M. & Artinger F. & Kämmer, J. E. & Gigerenzer G. (2014) "Big data" needs an analysis of decision processes, Journal of The Behavioral and brain sciences, 37. Cambridge, New York: Cambridge University Press;
- Andersson, G. (1988). Kritik und Wissenschaftsgeschichte. Kuhns, Lakatos' und Feyerabend's Kritik des Kritischen Rationalismus. Tübingen: Mohr;
- Borkar, V. & Carey, M. J. & Chen L. (2012) Inside "Big Data management": ogres, onions, or parfaits?, Proceedings of the 15th International Conference on Extending Database Technology, New York: ACM;
- Bentley, R.A. & O'Brien, M.J. & Brock W.A. (2014). Mapping collective behavior in the big-data era. Journal of The Behavioral and brain sciences, 37. Cambridge, New York: Cambridge University Press;
- Bookstein, F. L. (2014). The map is not the territory, Journal of The Behavioral and brain sciences, 37. Cambridge, New York: Cambridge University Press;
- Brendan, V. A., Verdoodt, V., Heyman, R., Ausloos, J., Wauters, E., Acar, G. (2015). From social media service to advertising network: A critical analysis of Facebook's Revised Policies and Terms - Provisional Report, Draft 23. Leuven: KU Leuven;
- Cavoukian, A. & Jonas, J. (2012). Privacy by design in the age of big data, Victoria: Office of the Information and Privacy Commissioner;

- Crawford, K. & Boyd, D. (2012). Critical Questions for Big Data - Provocations for a cultural, technological, and scholarly phenomenon, *Information, Communication & Society*, 15, London: Routledge
- Dhar, V. (2013). Data science and prediction, *Communications of the ACM CACM*, 56/12, New York: ACM;
- Douglas, L. & Beyer M. A. (2012). The Importance of 'Big Data': A Definition", Stamford: Gartner Group;
- Ekbja, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V. R., Tsou, A., Weingart, S. and Sugimoto, C. R. (2014), Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology*, 66(5). New York: John Wiley & Sons;
- Franks, B. (2012). Taming The Big Data Tidal Wave: Finding Opportunities in Huge Data Streams, Oxford, NJ: Wiley & Sons;
- Fuchs, C. (2013). Class and exploitation on the Internet, Digital labor, The Internet as playground and factory, Scholz, T. (Hg.), New York: Routledge.
- Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In Wright, G. & Ayton, P. (Hg.) Subjective probability. Oxford: Wiley & Sons;
- Gigerenzer, G. (1996). On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky. *Psychological Review*, 103, Washington: American Psychological Association;
- Gradinarua, A. (2014). The Contribution of Behavioral Economics in Explaining the Decisional Process, *Procedia Economics and Finance*, 16, Ubk.: Elsevier

- Herzog, W. (2012). Wissenschaftstheoretische Grundlagen der Psychologie, Kriz, J. (Hg), Osnabrück: Springer;
- Jacobs, A. (2009). The pathologies of big data, *Communications of the ACM - A Blind Person's Interaction with Technology*, 52/1, New York: ACM;
- Kahneman, D. & Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree, *The American Psychologist* 2009, 64, Washington: American Psychological Association;
- Kahneman D. (2011a). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux;
- Kallinikos, J. & Ekbja, H. & Nardi, B. (2015). Regimes of Information and the Paradox of Embeddedness, *The Information Society*, 31/2, New York: Taylor and Francis
- Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*, Second Edition. Chicago, London: The University of Chicago Press, Ltd;
- Kuhn, T. S. (1974). *The Essential Tension: Selected Studies in Scientific Tradition and Change*, Second Thoughts on paradigms, Chicago: University of Chicago Press;
- Kuhn, T. S. (1977). Objectivity, Value Judgment, and Theory Choice. *The British Journal for the Philosophy of Science*, 24. Oxford: British Society for the Philosophy of Science;
- Labrecque, L. I. & Vor dem Esche, J. & Mathwick, C. & Novak, T. P. & Hofacker, C. F. (2013). Consumer Power: Evolution in the Digital Age, *Journal of Interactive Marketing*, 27, Ubk.: Elsevier;

- Lazer, D. & Kennedy, R. & King, G. & Vespignani, A. (2013) The Parable of Google Flu: Traps in Big Data Analysis. *Science* 14, 343, Washington: Science/AAAS;
- Matthews, M. R. (2003). Thomas Kuhn's impact on science education: What lessons can be learned?, *Science Education*, 88/1, Ubk.: Wiley Periodicals;
- Marcinkowski, M. & Fonseca, F. (2015). The conditions of peak empiricism in big data and interaction design, *Journal of the Association for Information Science and Technology*, 66, Ubk.:ASIS&T;
- Mayer-Schönberger, V. & Cukier K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think* [Kindle Edition], Luxemburg: Amazon Digital Services, Inc.;
- Mayer-Schönberger, V. (2010) *Delete: The Virtue of Forgetting in the Digital Age*. Princeton, NJ: Princeton University Press;
- Ohm, P. (2010). *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*. *UCLA Law Review*, 57. Los Angeles: UCLA School of Law
- Popper, K. (1935). *Logik der Forschung - Zur Erkenntnistheorie der modernen Naturwissenschaft*, Wien: Julius Springer;
- Popper, K. (1972). *Die Logik der Sozialwissenschaften*. In Adorno et al., *Der Positivismusstreit in der deutschen Soziologie*. Darmstadt: Hermann Luchterhand.
- Popper, K. (1972). *Was ist Dialektik?* In Topitsch, E. (Hg.) (1972): *Logik der Sozialwissenschaften* (8. Aufl.). Köln: Kiepenheuer & Witsch.
- Schwab, K. & Marcus, A. & Oyola, J. R. & Hoffman W. (2011) *Personal Data: The Emergence of a New Asset Class*, World Economic Forum 2001;

Snijders, C. & Matzat, U. & Reips U. D. (2012). "Big Data": Big Gaps of Knowledge in the Field of Internet Science, International Journal of Internet Science, 7, Ubk.: Ijis

Squazzoni, F. & Jager, W. & Edmonds, B. (2013). Social Simulation in the Social Sciences: A Brief Overview, Social Science Computer Review, 32/3, London: SAGE Publishing;

Onlinequellen

- Ariely, D. (2013). Big Data is [...],
<https://twitter.com/danariely/status/287952257926971392> [Letzter Aufruf: 19. 04. 2015];
- Biselli A. (2015). „It’s a bug, not a feature“ – Sagt Facebook zum Tracking von Nicht-Usern... , <https://netzpolitik.org/2015/its-a-bug-not-a-feature-sagt-facebook-zum-tracking-von-nicht-usern/> [Letzter Aufruf: 11. 04. 2015];
- Butler, D. (2013). When Google got flu wrong, <http://www.nature.com/news/when-google-got-flu-wrong-1.12413> [Letzter Aufruf: 19. 04. 2015];
- Crawford, K. (2013). The Hidden Biases in Big Data, Harvard Business Review, <https://hbr.org/2013/04/the-hidden-biases-in-big-data> [Letzter Aufruf: 19. 04. 2015];
- Cukier, N. & Mayer Schönberger, V. (2013). The Rise of Big Data - How It's Changing the Way We Think About the World, <http://www.foreignaffairs.com/articles/139104/kenneth-neil-cukier-and-viktor-mayer-schoenberger/the-rise-of-big-data> [Letzter Aufruf: 19. 04. 2015];
- Demgen, A. (2015). Google: Diese Kriterien entscheiden über Löschung eines Links, <http://www.netzwelt.de/news/151080-google-kriterien-entscheiden-ueber-loeschung-links.html> [Letzter Aufruf: 19. 04. 2015];
- Duhigg, C. (2012). How Companies Learn Your Secrets, http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=0 [Letzter Aufruf: 19. 04. 2015];
- Fairless, T. (2015). Europe’s Digital Czar Slams Google, Facebook. Wall Street Journal, <http://www.wsj.com/articles/europes-digital-czar-slams-google->

facebook-over-selling-personal-data-1424789664 [Letzter Aufruf: 24. 02. 2015];

Gibbs S. (2015). Facebook's privacy policy breaches European law, report finds, <http://www.theguardian.com/technology/2015/feb/23/facebooks-privacy-policy-breaches-european-law-report-finds> [Letzter Aufruf: 11. 04. 2015];

Haddow, D. (2014). DATAcide - The Total Annihilation of Life as We Know It, <https://www.adbusters.org/magazine/115/datacide-total-annihilation-life-we-know-it.html> [Letzter Aufruf: 19. 04. 2015];

Haddow, D. (2015). Life in the Algorithm, <https://www.adbusters.org/magazine/117/life-algorithm.html> [Letzter Aufruf: 19. 04. 2015];

Harford, T. (2014). Big data: are we making a big mistake?, <http://www.ft.com/intl/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html> [Letzter Aufruf: 26. 02. 2015];

Kahneman, D. (2011b). @Google Presents: Daniel Kahneman, Talks at Google, <https://www.youtube.com/watch?v=CjVQJdlrDJ0> [Letzter Aufruf: 19. 04. 2015];

Klimburg, A. (2014). Einleitende Rede zum Vortrag - Big Data: Ein Bergwerk ohne Regeln?, <http://www.alpbach.org/de/vortrag/big-data-ein-bergwerk-ohne-regeln/> [Letzter Aufruf: 11.04.2015];

Kordik, H. (2009). Josef Prölls funkelnagelneue Denkfabrik, <http://diepresse.com/home/wirtschaft/economist/kordiconomy/477669/> [Letzter Aufruf: 19. 04. 2015];

Krempl, S. (2015). EU-Staaten verabschieden sich von der Datensparsamkeit, <http://www.heise.de/newsticker/meldung/EU-Staaten-verabschieden-sich->

von-der-Datensparsamkeit-2574967.html [Letzter Aufruf: 19. 04. 2015];

Laney, D. (2001). 3-D Data Management: Controlling Data Volume, Velocity and Variety, <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> [Letzter Aufruf: 26.02.2015];

Laney, D. (2012). Deja VVVu: Others Claiming Gartner's Construct for Big Data, <http://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/> [Letzter Aufruf: 26.02.2015];

Möchel, E. (2013). EU-Ministerrat demontiert Datenschutzpaket, <http://fm4.orf.at/stories/1718534/> [Letzter Aufruf: 19. 04. 2015];

Naughton J. (2012). Thomas Kuhn: the man who changed the way the world looked at science, <http://www.theguardian.com/science/2012/aug/19/thomas-kuhn-structure-scientific-revolutions> [Letzter Aufruf: 19. 04. 2015];

O'Neill, P. H. (2015). Former Chief Privacy Officer warned Microsoft of NSA in 2011, <http://www.dailydot.com/politics/microsoft-prism-2011-fired/> [Letzter Aufruf 19.04.2015];

Rath, C. (2014). Das Recht auf Privatheit überwiegt, <http://www.taz.de/1/archiv/?dig=2014/09/19/a0084> [Letzter Aufruf: 24. 02. 2015];

Ubk. FAZ (2014). So können Google-Nutzer Suchergebnisse löschen lassen, <http://www.faz.net/aktuell/wirtschaft/netzwirtschaft/neues-formular-im-internet-so-koennen-google-nutzer-suchergebnisse-loeschen-lassen-12964383.html> [Letzter Aufruf: 11.04.2015];

Ubk. KU Leuven (2015). ICRI/CIR and iMinds-SMIT advise Belgian Privacy Commission in Facebook investigation,

<http://www.law.kuleuven.be/icri/en/news/item/icri-cir-advises-belgian-privacy-commission-in-facebook-investigation> [Letzter Aufruf: 19. 04. 2015];

Ubk. re:publica (2014). RE:PUBLICA 2014 - Viktor Mayer-Schönberger - Freiheit und Vorhersage, <http://re-publica.de/file/republica-2014-viktor-mayer-schoenberger-freiheit-un> [Letzter Aufruf: 19.04.2015];

Ubk., (2014), Biography, <http://timharford.com/etc/biography/> [Letzter Aufruf: 19. 04. 2015];

Ubk., (2015), Facebook hat alle Daten aus Gesichtserkennung gelöscht .
<http://www.zeit.de/digital/datenschutz/2013-02/facebook-gesichtserkennung-verfahren-caspar> [Letzer Aufruf: 20. 04. 2015]

Ubk., (2015). Wikipedia Eintrag - Recht auf Vergessenwerten,
http://de.wikipedia.org/wiki/Recht_auf_Vergessenwerden , [Letzter Aufruf: 19.04.2015];

Ubk., Alpbach Talks (2013). Themenreihe „Digitale Welten“: Der digitale Mensch im Web 3.0, <http://www.alpbach.org/de/unterveranstaltung/alpbach-talks-big-data/> [Letzter Aufruf: 23. 02. 2015];

Ubk., Alpbach Talks (2014). Big Data: Ein Bergwerk ohne Regeln?,
<http://www.alpbach.org/de/vortrag/big-data-ein-bergwerk-ohne-regeln/>
[Letzter Aufruf: 23. 02. 2015];

Ubk., APA (2014). Alpbacher Gesundheitsgespräche: Keine Zukunft ohne “Big Data”,
<https://medonline.at/2014/europaeisches-forum-alpbach-keine-zukunft-ohne-big-data/> [Letzter Aufruf: 11.04.2015];

Ubk., APA (2015). "Lösch-Beirat" von Google uneins über Recht auf Vergessen,
<http://futurezone.at/netzpolitik/loesch-beirat-von-google-uneins-ueber-recht-auf-vergessen/112.272.815> [Letzter Aufruf: 24. 02. 2015];

Ubk., Der Spiegel (2015). Bundesnachrichtendienst: Bundesregierung plant Gesetz zur Spionage im Weltraum,
<http://www.spiegel.de/netzwelt/netzpolitik/bundesregierung-plant-gesetz-zur-spionage-im-weltraum-a-1027963.html> [Letzter Aufruf: 19. 04. 2015];

Ubk., Der Standard (2013). 3.133 Änderungsanträge für EU-Datenschutz-Reform,
<http://derstandard.at/1362108027424/3133-Aenderungsantraege-fuer-EU-Datenschutz-Reform> [Letzter Aufruf: 19. 04. 2015];

Ubk., Südwest Presse (2010). Mehr als 1500 Änderungsanträge,
<http://www.swp.de/ulm/nachrichten/politik/Mehr-als-1500-Aenderungsantraege;art4306,406656> [Letzter Aufruf: 19. 04. 2015];

Williams, B. (2014). NBC News, <http://www.nbcnews.com/feature/edward-snowden-interview/edward-snowden-timeline-n114871> [Letzter Aufruf: 23. 02. 2015];

Zimmer, M. (2008). More On the “Anonymity” of the Facebook Dataset – It's Harvard College (Updated),<http://www.michaelzimmer.org/2008/10/03/more-on-the-anonymity-of-the-facebook-dataset-its-harvard-college/> [Letzter Aufruf: 19. 04. 2015];

Abbildungsverzeichnis

Abbildung 1 – Volltextsuche nach „Snowden“; selbst produziert

Abbildung 2 – Big Data Metrik: Volume, Velocity, Variety, nach Laney; Aus: Laney, D. (2012). Deja VVVu: Others Claiming Gartner's Construct for Big Data, <http://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/> [Letzter Aufruf: 26.02.2015];

Abbildung 3 – „Generalisierte Verteilungen die die unterschiedlichen Quadranten der Karte charakterisieren. Normal (Gauss) im Nordwesten, negativ binominal im Südwesten, log-normal im Nordosten und power-law im Südosten“; Aus: Bentley et al. (2014). Mapping collective behavior in the big-data era. Journal of The Behavioral and brain sciences, 37. Cambridge, New York: Cambridge University Press;

Abbildung 4 – „Zusammenfassung der Vier-Quadranten-Karte zum Verständnis unterschiedlicher Felder des Entscheidungsprozesses von Personen, darauf basierend ob eine Entscheidung unabhängig, oder sozial beeinflusst getroffen wird, sowie der Transparenz der Entscheidungsoptionen sowie ihres jeweiligen „Payoffs“; Aus: Bentley et al. (2014). Mapping collective behavior in the big-data era. Journal of The Behavioral and brain sciences, 37. Cambridge, New York: Cambridge University Press;

Abbildung 5 - Einleitung der Sektion 4 des neuen Entwurfs der Datenschutzdirektive. Inhaltlich veränderte Stelle durch Unterstreichen hervorgehoben.; Aus: Möchel, E. (2013). EU-Ministerrat demontiert Datenschutzpaket, <http://fm4.orf.at/stories/1718534/> [Letzter Aufruf: 19. 04. 2015];

Wissenschaftlicher Lebenslauf

Persönliche Daten

Name: Philipp Ganster, Bakk.
Geburtsdatum: 09. 03. 1984
Anschrift: Schloss Schönbrunn,
Meidlinger Viereckl 116
1130 Wien
E-Mail: ganster.philipp@gmail.com

Juni 2002 - Abschluss am BG/BRG 13, Fichtnergasse, Matura

Oktober 2003 - Beginn des Bakkalaureatsstudiums der Publizistik- und Kommunikationswissenschaften an der Universität Wien.

Studienschwerpunkte: Print-Journalismus, Online-Journalismus und kommunikationswissenschaftliche Forschung;

November 2008 – Beginn des Magisterstudiums der Publizistik- und Kommunikationswissenschaften an der Universität Wien.

Studienschwerpunkte: Neue Medien, Public Relations, Gruppenkommunikation;

Studienbegleitend konnte ich in dieser Zeit Erfahrungen als freier Journalist in Jugendmagazinen, sowie als Praktikant in der Kulturabteilung (FP4) des ORF, sowie in der Marketingabteilung von Microsoft Österreich (Entertainment and Devices) sammeln. Einen ersten Kontakt zum Feld der automatisierten Informationsbearbeitung erarbeitete ich mir in der Konzeption und Erstellung des wichtigsten deutschen

Szeneleitfadens zur Digitalisierung von Büchern, zum Zwecke ihrer Überführung in ein eReader-taugliches Format.

Abstract

Deutsch

Big Data Anwendungen betreffen unser tägliches Leben immer häufiger und finden als nur mehr schwer zu ignorierender Ansatz auch verstärkt Beachtung in sozial- und geisteswissenschaftlichen Forschungsrichtungen.

Diese Arbeit sucht Antworten auf Fragen nach der gesellschaftlichen Relevanz und der aktuellen Bedeutung des Big Data Ansatzes und versucht seine Vereinbarkeit mit dem sozialwissenschaftlichen Wissenschaftsverständnis zu klären.

Dazu beinhaltet sie eine Analyse der aktuellen interdisziplinären Diskussion, des gesellschaftspolitischen Spektrums rund um das Themenfeld, eine Prüfung der wissenschaftstheoretischen Vereinbarkeit mit dem post-positivistischen Wissenschaftsverständnis, sowie eine inhaltliche Analyse der davon berührten Entscheidungstheorien aus anderen Forschungsrichtungen.

Im letzten Teil wird zur Ethik und den gesellschaftlichen Auswirkungen des Big Data Ansatzes geforscht.

English

Big Data applications increasingly impact our daily life – and as an approach it is increasingly hard to ignore in the fields of social sciences and the humanities. This work seeks to answer questions about the current significance and the social relevance of the Big Data approach and attempts to clarify its compatibility with the sociological understanding of science. This dissertation includes an analysis of the current interdisciplinary discussion, of the socio-political spectrum impacted by the Big Data approach, an examination of the scientific compatibility within the paradigm of postpositivism - and a content analysis as it touches decision making theories from other fields of research. The last part of the paper touches on ethics and the social impact of the Big Data approach.